

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 May 2001 (17.05.2001)

PCT

(10) International Publication Number
WO 01/35317 A1

- (51) International Patent Classification⁷: **G06F 19/00** (74) Agent: **EINHORN, Gregory, P.**; Fish & Richardson P.C., Suite 500, 4350 La Jolla Village Drive, San Diego, CA 92122 (US).
- (21) International Application Number: **PCT/US00/31152**
- (22) International Filing Date:
13 November 2000 (13.11.2000)
- (25) Filing Language: **English**
- (26) Publication Language: **English**
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (30) Priority Data:
60/165,086 12 November 1999 (12.11.1999) US
60/165,124 12 November 1999 (12.11.1999) US
60/179,531 1 February 2000 (01.02.2000) US
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant (*for all designated States except US*): **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 12th floor, 1111 Franklin Street, Oakland, CA 94607-5200 (US).
- Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*
- (72) Inventors; and
(75) Inventors/Applicants (*for US only*): **EISENBERG, David** [US/US]; 342 Comstock Avenue, Los Angeles, CA 90024 (US). **ROTSSTEIN, Sergio, H.** [—/US]; — (US). **MARCOTTE, Edward, M.** [US/US]; 2230 S. Carmelina Avenue, Los Angeles, CA 90064 (US).
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: **DETERMINING THE FUNCTIONS AND INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS**

(57) Abstract: The invention provides novel methods for characterizing the function of nucleic acids and polypeptides. The invention provides a novel method for identifying a nucleic acid or a polypeptide sequence that may be a target for a drug. The invention provides a novel method for identifying a nucleic acid or a polypeptide sequence that may be essential for the growth or viability of an organism. The characterization is based on use of methods of the invention comprising algorithms that can identify functional relationships between diverse sets of non-homologous nucleic acid and polypeptide sequences. The invention provides a computer program product, stored on a computer-readable medium, for identifying a nucleic acid or a polypeptide sequence that may be essential for the growth or viability of an organism. The invention provides a computer program product, stored on a computer-readable medium, for identifying a nucleic acid or a polypeptide sequence that may be a target for a drug. The invention provides a computer system, comprising a processor and a computer program product of the invention.

WO 01/35317 A1

DETERMINING THE FUNCTIONS AND INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS

Related Applications

The present application is a continuation-in-part application ("CIP") of Patent
5 Convention Treaty (PCT) International Application Serial No: PCT/US00/02246, filed in the
U.S. receiving office on January 28, 2000, and this application claims the benefit of priority
under 35 U.S.C. § 119(e) of U.S. Provisional Application Nos. 60/165,124, and 60/165,086,
both filed November 12, 1999, and U.S. Provisional Application No. 60/179,531, filed February
1, 2000. International Application Serial No: PCT/US00/02246 claims the benefit of priority
10 under 35 U.S.C. § 119(e) of U.S. Provisional Application Serial No. 60/117,844, filed January
29, 1999, U.S. Provisional Application Serial No. 60/118,206, filed February 1, 1999, U.S.
Provisional Application Serial No. 60/126,593, filed March 26, 1999, U.S. Provisional
Applications Serial No. 60/134,093, filed May 14, 1999, and U.S. Provisional Application
Serial No. 60/134,092, filed May 14, 1999. Each of the aforementioned applications is
15 explicitly incorporated herein by reference in their entirety and for all purposes.

TECHNICAL FIELD

This invention generally relates to genetics and microbiology. The invention
provides novel methods to identify the function of and relationships between nucleic acid and
protein sequences. The method is particularly useful for finding the identifying genes and
20 polypeptides having potential therapeutic relevance in organisms, e.g., microorganisms, such
as *Mycobacterium tuberculosis*. The invention also provides *Mycobacterium tuberculosis*
genes and polypeptides found by these methods. These genes and polypeptides are useful as
potential drug targets.

BACKGROUND

25 The determination of the functions of and relationships between nucleic acid
and protein sequences has traditionally relied on either the study of homology and sequence
identity with genes and proteins of known function or, in the absence of informative
homology, laborious experimental work. The availability of many complete genome
sequences has made it possible to develop new strategies for computational determination of
30 protein functions. Several methods have been developed which can predict the general

function of proteins by analyzing their functional relationships rather than sequence similarity. Generally, two proteins can be considered functionally related when they form part of the same biochemical pathway or biological process. For example, although malate dehydrogenase is not homologous to pyruvate carboxylase, and the two enzymes do not
5 catalyze the same reaction, they are functionally related because they both catalyze steps of a common biochemical pathway, namely the tricarboxylic acid cycle.

New methods that can establish such functional relationships could provide valuable information on the functions of uncharacterized nucleic acid and protein sequences.

The disease tuberculosis, caused *Mycobacterium tuberculosis* (MTB) is one
10 of the world's leading killers. The World Health Organization estimates that 30 million deaths from pulmonary tuberculosis will occur during this decade. Alarming reports on the emergence of drug-resistant strains of this bacterium underscore the importance of the search for new therapeutic agents. Identifying the function of every protein produced by MTB will provide researchers with promising new targets for anti-tuberculosis drug design.

15 SUMMARY

The invention provides novel methods for characterizing the function of nucleic acids and polypeptides. The invention provides a novel method for identifying a nucleic acid or a polypeptide sequence that may be a target for a drug. The invention provides a novel method for identifying a nucleic acid or a polypeptide sequence that may be essential
20 for the growth or viability of an organism. The characterization is based on use of methods of the invention comprising algorithms that can identify functional relationships between diverse sets of non-homologous nucleic acid and polypeptide sequences. Characterization of nucleic acid and protein sequences can be the basis for the development of compositions that can interact with those nucleic acids and polypeptides. For example, such characterization can
25 provide a basis for screening methods. Such characterization may allow use of these sequences as targets for drug discovery. Discovery of such compositions can provide the basis for the design of novel drugs, particularly if the characterized sequences are derived from a pathogen.

The invention provides a method for identifying a nucleic acid or a
30 polypeptide sequence that may be a target for a drug comprising the following steps: (a)

providing a first nucleic acid or a polypeptide sequence that is known to be a drug target; (b) providing at least one algorithm selected from the group consisting of a "domain fusion" method, a "phylogenetic profile" method and a "physiologic linkage" method, wherein the algorithm is capable analyzing a functional relationship between nucleic acid or polypeptide sequences; and, (c) comparing the first nucleic acid or the polypeptide drug target sequence to a plurality of sequences using at least one of the algorithms as set forth in step (b) to identify a second sequence that has a functional relationship to the first sequence, thereby identifying a nucleic acid or a polypeptide sequence that may be a target for a drug.

The invention provides a method for identifying a nucleic acid or a polypeptide sequence that may be essential for the growth or viability of an organism comprising the following steps: (a) providing a first nucleic acid or a polypeptide sequence that is known to be essential for the growth or viability of an organism; (b) providing at least one algorithm capable analyzing a functional relationship between nucleic acid or polypeptide sequences selected from the group consisting of a "domain fusion" method, a "phylogenetic profile" method and a "physiologic linkage" method; and, (c) comparing the first nucleic acid or the polypeptide sequence to a plurality of sequences using at least one of the algorithms as set forth in step (b) to identify a second sequence that has a functional relationship to the first sequence, thereby identifying a nucleic acid or a polypeptide sequence that may be essential for the growth or viability of an organism.

In one aspect of the methods of the invention, the drug is an anti-microbial drug. In another aspect, the first nucleic acid or a polypeptide sequence is derived from a pathogen. The pathogen can be a microorganism, such as *Mycobacterium tuberculosis* (MTB).

The plurality of sequences used to identify a second sequence can comprise a database of the gene sequences of an entire genome of an organism. The plurality of sequences used to identify a second sequence can comprise a database of the gene sequences derived from a pathogen.

In one aspect of the methods of the invention, the "phylogenetic profile" method algorithm comprises (a) obtaining data, comprising a list of proteins from at least two genomes; (b) comparing the list of proteins to form a protein phylogenetic profile for each protein, wherein the protein phylogenetic profile indicates the presence or absence of a

protein belonging to a particular protein family in each of the at least two genomes based on homology of the proteins; and (c) grouping the list of proteins based on similar profiles, wherein proteins with similar profiles are indicated to have a functional relationship. The phylogenetic profile can be in the form of a vector, matrix or phylogenetic tree. The

5 "phylogenetic profile" method can further comprise determining the significance of homology between the proteins by computing a probability (p) value threshold. The probability can be set with respect to the value $1/NM$, based on the total number of sequence comparisons that are to be performed, wherein N is the number of proteins in the first organism's genome and M in all other genomes. The presence or absence of a protein

10 belonging to a particular protein family in each of the at least two genomes can be determined by calculating an evolutionary distance. The evolutionary distance can be calculated by: (a) aligning two sequences from the list of proteins; (b) determining an evolution probability process by constructing a conditional probability matrix: $p(aa \rightarrow aa')$, where aa and aa' are any amino acids, said conditional probability matrix being constructed

15 by converting an amino acid substitution matrix from a log odds matrix to said conditional probability matrix; (c) accounting for an observed alignment of the constructed conditional probability matrix by taking the product of the conditional probabilities for each aligned pair during the alignment of the two sequences, represented by $P(p) = \prod_n p(aa_n \rightarrow aa'_n)$; and, (d)

determining an evolutionary distance α from powers equation $p' = p^\alpha(aa \rightarrow aa')$, maximizing

20 for P . The conditional probability matrix can be defined by a Markov process with substitution rates, over a fixed time interval. The conversion from an amino acid substitution matrix to a conditional probability matrix can be represented by:

$$P_B(i \rightarrow j) = p(j) 2^{\frac{\text{BLOSUM62}_{ij}}{2}},$$

where BLOSUM62 is an amino acid substitution matrix, and $P(i \rightarrow j)$ is the

25 probability that amino acid i is replaced by amino acid j through point mutations according to BLOSUM62 scores. In one aspect, the P_j 's are the abundances of amino acid j and are computed by solving a plurality of linear equations given by the normalization condition that:

$$\sum_i P_B(i \rightarrow j) = 1.$$

In alternative aspects of the methods of the invention, the “physiologic linkage” method algorithm identifies proteins and nucleic acids that participate in a common functional pathway; identifies proteins and nucleic acids that participate in the synthesis of a common structural complex; and, identifies proteins and nucleic acids that participate in a common metabolic pathway.

In one aspect of the invention, the “domain fusion” method algorithm comprises (a) aligning a first primary amino acid sequence of multiple distinct non-homologous polypeptides to second primary amino acid sequence of a plurality of proteins; and, (b) for any alignment found between the first primary amino acid sequences of all of such multiple distinct non-homologous polypeptides and at least one protein of the second primary amino acid sequences, outputting an indication identifying the aligned second primary amino acid sequence as an indication of a functional link between the aligned first and second polypeptide sequences. The aligning can be performed by an algorithm selected from the group consisting of a Smith-Waterman algorithm, Needleman-Wunsch algorithm, a BLAST algorithm, a FASTA algorithm, and a PSI-BLAST algorithm. The multiple distinct non-homologous polypeptides can be obtained by translating a nucleic acid sequence from a genome database. The plurality of proteins can have a known function. At least one of the multiple distinct non-homologous polypeptides can have a known function. At least one of the multiple distinct non-homologous polypeptides can have an unknown function. The alignment can be based on the degree of homology of the multiple distinct non-homologous polypeptides to the plurality of proteins. The “domain fusion” method can comprise determining the significance of the aligned and identified second primary amino acid sequence by computing a probability (p) value threshold. The probability threshold can be set with respect to the value $1/NM$, based on the total number of sequence comparisons that are to be performed, wherein N is the number of proteins in a first organism’s genome and M in all other genomes. The “domain fusion” method can further comprising filtering excessive functional links between one first primary amino acid sequence of multiple distinct non-homologous polypeptides and an excessive number of other distinct non-homologous polypeptides for any alignment found between the first primary amino acid sequences of the

distinct non-homologous polypeptides and at least one of the second primary amino acid sequences of the plurality of proteins.

The invention provides a computer program product, stored on a computer-readable medium, for identifying a nucleic acid or a polypeptide sequence that may be a target for a drug, the computer program product comprising instructions for causing a computer system to be capable of: (a) inputting a first nucleic acid or a polypeptide sequence that is known to be a drug target; (b) accessing at least one algorithm capable analyzing a functional relationship between nucleic acid or polypeptide sequences selected from the group consisting of a "domain fusion" method, a "phylogenetic profile" method and a "physiologic linkage" method; and (c) comparing the first nucleic acid or the polypeptide drug target sequence to a plurality of sequences using at least one of the algorithms set forth in step (b) to identify a second sequence that has a functional relationship to the first sequence and generating an output identifying a nucleic acid or a polypeptide sequence that may be a target for a drug .

The invention provides a computer program product, stored on a computer-readable medium, for identifying a nucleic acid or a polypeptide sequence that may be essential for the growth or viability of an organism, the computer program product comprising instructions for causing a computer system to be capable of: (a) providing a first nucleic acid or a polypeptide sequence that is known to be essential for the growth or viability of an organism; (b) accessing at least one algorithm capable analyzing a functional relationship between nucleic acid or polypeptide sequences selected from the group consisting of a "domain fusion" method, a "phylogenetic profile" method and a "physiologic linkage" method; and, (c) comparing the first nucleic acid or the polypeptide sequence to a plurality of sequences using at least one of the algorithms set forth in step (b) to identify a second sequence that has a functional relationship to the first sequence and generating an output identifying a nucleic acid or a polypeptide sequence that may be essential for the growth or viability of an organism.

The invention provides a computer system, comprising: (a) a processor; and, a computer program product of the invention.

All publications, patents, patent applications, GenBank sequences and ATCC deposits, cited herein are hereby expressly incorporated by reference for all purposes.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

Figure 1 is an example of functional linkages predicted between InhA (Rv 1484) and other TB genes.

Figure 2 is an example of predicted functional linkages between embB (Rv 3795), which is a target of the drug ethambutol, and other TB genes using the phylogenetic profile method.

Figure 3 is an example of predicted functional linkages between five TB genes having homology to penicillin binding proteins and other TB genes.

Figure shows that gcpE (Rv 2868C) is predicted to be functional linked to cell wall metabolism.

Figure 5 shows predicted functional linkages of htrA (Rv 1223C) with other TB genes.

Like reference symbols in the various drawings indicate like elements.

DETAILED DESCRIPTION

The present invention provides novel methods for identifying the relationships between and the function of nucleic acid and polypeptide sequences. The methods of the invention identify novel genes and polypeptides on the basis of their functional linkage to other proteins whose biological function or processes is known or inferred by homology.

The genes and polypeptides identified by the methods of the invention can be used in screening methods for the identification of compositions which, by binding or otherwise interacting with the gene or polypeptide, are capable of modifying the physiology and growth of an organism. The compositions identified by these screening methods are useful as drugs and pharmaceuticals. Thus, genes and polypeptides identified by the methods

of the invention, including the genes and polypeptides identified herein, can be used as potential drug targets.

One aspect of the invention provides methods for identifying the function of genes and polypeptides from *Mycobacterium tuberculosis* (MTB or TB). Based on this new functional determination, these genes and polypeptides can be used to screen for compositions capable of modifying the physiology and growth of *Mycobacterium tuberculosis* (TB). Thus, genes and polypeptides identified by the methods of the invention, including the genes and polypeptides identified herein, can be used as targets in screening protocols and can be useful as potential drug targets.

The function of the TB genes and polypeptides of the present invention were identified using the methods of the invention; i.e., they were identified on the basis of their functional linkage to other proteins whose biological function or processes were known by experiment or inferred by homology. TB genes and polypeptides that are functionally linked to genes known to be involved in pathogenesis or organisms survival are potential drug targets. Genes or polypeptides associated with TB pathogenesis, survival or that are important or unique to TB biochemical pathways are potential drug targets. TB genes and polypeptides that have no homologues identified in humans are potential drug targets. The function of many of the TB genes and polypeptides identified is based on the genes or polypeptides with which they are functionally linked.

TB genes whose function was identified using the methods of the invention are effectively targeted by a drug (i.e., they can act as *bona fide* drug targets) provides proof of principle that the invention's methods for identifying functionally linked genes can identify TB genes and polypeptides that are drug targets. Further confirmation that the genes identified by the methods of the invention include *bona fide* drug targets can be supported by the fact that genes already known to be targets for drugs have been independently identified, or "re-discovered," by the invention's methods.

The novel TB genes described herein are identified as being functionally related or linked to other genes, including other TB genes, such as a known TB drug target (e.g., InhA polypeptide, which is a target of isoniazid). These functional linkages are established using mathematical algorithms. The assignment or inference of a function to TB genes and polypeptides based on their linkage or relatedness to other genes and polypeptides

is described in U.S. provisional application serial no. 60/165,086. Potential TB drug targets are identified by several methods discussed herein and in further detail in U.S. provisional application serial no. 60/134,092. Through the use of these methods, TB genes and polypeptides have been identified as potential drug targets and are illustrated on Tables 1 and 2, and Figures 1 to 5. The nucleotide and amino acid sequences of these potential drug targets are illustrated on Tables 3 and 4, respectively (see below).

The phrase "functional link," "functionally related" and grammatical variations thereof, when used in reference to genes or polypeptides, means that the genes or polypeptides are predicted to be linked or related. A particular example of functionally related or linked proteins is where two proteins participate in a biochemical or metabolic pathway (e.g., malate dehydrogenase and fumarase, which are both present in the TCA cycle). Thus, although functionally linked or related proteins may not have sequence homology to each other, they are linked by virtue of their participation in the same biochemical pathway. Other examples of linked or related polypeptides are where two polypeptides are part of a protein complex, physically interact, or act upon each another.

The "domain fusion" or "Rosetta Stone" method searches protein sequences across all known genomes and identifies proteins that are separate in one organism but joined as intramolecular domains into one larger protein in another organism. Such proteins that are separate in some organisms but joined in others often carry out related or sequential functions and are therefore functionally linked.

The phylogenetic profile method compares protein sequences across all known genomes and analyzes the pattern of inheritance of each protein across the different organisms. Proteins that have similar patterns of inheritance, either acquired or lost as a part of a group of proteins through evolution, are functionally linked. The gene proximity method identifies genes that remain physically close or "clustered" throughout evolution and are therefore functionally linked.

A particular example of the identification of a potential TB drug target would be to identify a TB gene or polypeptide functionally linked to a known drug target. Anti-TB drugs include isoniazid, rifampicin, ethambutol, streptomycin, pyrazinamide, and thiacetazone. For isoniazid, this drug is believed to act through enoyl-acyl reductase InhA, resulting in mycolic acid biosynthesis inhibition. Thus, TB genes or polypeptides

functionally linked to enoyl-acyl reductase *InhA* are potential drug targets; see Figure 1, which shows an analysis of *InhA*, the target for isoniazid, the most widely used anti-tuberculosis drug, and functional linkages to a set of genes mostly known or hypothesized to be involved in cell wall-related processes and lipid and polyketide metabolism. Particular
5 examples of the identification of several TB genes and polypeptides that are functionally related to the target of these anti-TB drugs is shown in Figures 1 to 5.

“Domain Fusion” or “Rosetta Stone” Method

The “domain fusion” or “Rosetta Stone” method compares protein sequences across known nucleic acid databases (e.g., known genomes) to identify genes and proteins
10 that are separate entities in one organism but are joined into one larger multidomain protein in another organism. In such cases, the two separate proteins often carry out related or sequential functions or form part of a larger protein complex. Therefore, the general function of one component (e.g., one or more of the unknown proteins) can be inferred from the known function of the other component. In addition, merely identifying links between
15 proteins using the method described herein provides valuable information (e.g., usefulness as a target for an antibacterial drug), regardless of whether the function of one or more of the proteins used to form the link(s) is known. Because the two components do not have similar amino acid sequence the function of one could not be inferred from the other on the basis of sequence similarity alone.

20 The methods for identifying drug targets (e.g., TB drug targets) described herein (e.g., the “Rosetta Stone Method”) are based on the idea that proteins that participate in a common structural complex, metabolic pathway, biological process or with closely related physiological functions, are functionally linked. In addition, these methods also are capable of identifying proteins that interact physically with one another. Functionally linked
25 proteins in one organism can often be found fused into a single polypeptide chain in a different organism. Similarly, fused proteins in one organism can be found as individual proteins in other organisms. For example, in a first organism one might identify two unlinked proteins “A” and “B” with unknown function. In another organism, one may find a single protein “AB” with a part that resembles “A” and a part that resembles “B”. Protein
30 AB allows one to predict that “A” and “B” are functionally related.

The functional activity of each distinct protein in the "Rosetta Stone" method need not be known prior to performing the method (*i.e.*, the function of A, B, or AB need not be known). Using the "Rosetta Stone" method to compare and analyze several unknown protein sequences can provide information regarding relationships of each protein absent
5 knowledge about the functional activity of the initially analyzed proteins themselves. For example, the information (*i.e.*, the links) can provide information that the proteins are part of a common pathway, function in a related process or physically interact. Such information need not be based on the biological function of the individual proteins.

These methods can provide information regarding links between previously
10 un-linked proteins that function, for example, in a concerted process. A marker, for example, for a particular disease state is identified by the presence or absence of a protein (*e.g.*, Her2/neu in breast cancer detection). Links (*i.e.*, information) identified by the method, which link proteins "B" and "C" to such a marker suggest that proteins "B" and "C" are related by function, physical interaction or part of a common biological pathway with the
15 marker. Such information is useful in designing screening methods and identifying drug targets (*e.g.*, TB drug targets), making diagnostics, and designing therapeutics.

In one approach, the "Rosetta Stone" method is performed by sequence comparison that searches for incomplete "triangle relationships" between, for example, three proteins, *i.e.*, for two proteins A' and B' that are different from one another but similar in
20 sequence to another protein AB. Completing the triangle relationship provides useful information regarding the proteins' biological function(s), functional interaction, pathway relationships or physical relationships with other proteins in the "triangle."

Either nucleotide sequences or amino acid sequences can be used in the methods for identifying functionally related or linked genes or polypeptides. Where a
25 nucleic sequence is to be used it can be first translated from a nucleic acid sequence to amino acid sequence. Such translation may be performed in all frames if the coding sequence is not known. Programs that can translate a nucleic acid sequence are known in the art. In addition, for simplicity, the description of this method discusses the use of a "pair" of proteins in the determination of a "Rosetta Stone" protein, more than 2 may be used (*e.g.*, 3,
30 4, 5, 10, 100 or more proteins). Accordingly, one can analyze chains of linked proteins, such as "A" linked by a Rosetta Stone protein to "B" linked by a Rosetta Stone protein to "C", etc.

By this method, groups of functionally related proteins can be found and their function identified.

A method can start with identifying the primary amino acid sequence for a plurality of proteins whose functional relationship is to be determined (*e.g.*, protein A' and protein B'). A number of source databases are available, as described above, that contain
5 either a nucleic acid sequence and/or a deduced amino acid sequence for use with the first step. The plurality of sequences (the "probe sequences") are then used to search a sequence database, *e.g.*, GenBank (NCBI, NLM, NIH), PFAM (a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains;
10 Washington University, St. Louis MO) or ProDom (a database based on recursive PSI-BLAST searches and designed as a tool to help analyze domain arrangements of proteins and protein families, see, *e.g.*, Corpet (1999) *Nucleic Acids Res.* 27:263-267), either simultaneously or individually. Every protein in the sequence database is examined for its ability to act as a "Rosetta Stone" protein (*i.e.*, a single protein containing polypeptide
15 sequences or domains from both protein A' and protein B'). A number of different methods of performing such sequence searches are known in the art. Such sequence alignment methods include, for example, BLAST (see, *e.g.*, Altschul (1990) *J. Mol. Biol.* 215: 403-410), BLITZ (MPsrch) (see, *e.g.*, Brenner (1995) *Trends Genet.* 11:330-331; and *infra*), and FASTA (see, *e.g.*, Pearson (1988) *Proc. Natl. Acad. Sci. USA* 85(8):2444-2448; and *infra*).
20 The probe sequence can be any length (*e.g.*, about 50 amino acid residues to about 1000 amino acid residues).

Probe sequences (*e.g.*, polypeptide sequences or domains) found in a single protein (*e.g.*, an "AB" multidomain protein) are defined as being "linked" by that protein. Where the probe sequences are used individually to search the sequence database, one can
25 mask those segments having homology to the first probe sequence found in the proteins of the sequence database prior to searching with the subsequent probe sequence. In this way, one eliminates any potential overlapping sequences between the two or more probe sequences.

The linked proteins can then be further compared for similarity with one
30 another by amino acid sequence comparison. Where the sequences are identical or have high homology, such a finding can be indicative of the formation of homo-dimers, -trimers, etc.

Typically, "Rosetta Stone"-linked proteins are only kept when the linked proteins show no homology to one another (*e.g.*, hetero-dimers, trimers, etc.).

In another method for identifying functional linkages, a potential fusion protein lacking any functional information that is suspected of having two or more domains (*e.g.*, a potential "Rosetta Stone" protein) may be used to search for related proteins. In this method, the primary amino acid of the fusion protein is determined and used as a probe sequence. This probe sequence is used to search a sequence database (*e.g.*, GenBank, PFAM or ProDom). Every protein in the sequence database is examined for homology to the potential fusion protein (*i.e.*, multiple proteins containing polypeptide sequences or domains from the potential fusion protein). A number of different methods of performing such sequence searches are known in the art, *e.g.*, BLAST, BLITZ (Biocomputing Research Unit, University of Edinburgh, Scotland, the "MPsrch program" performs comparisons of protein sequences against the Swiss-Prot protein sequence database using the Smith and Waterman best local similarity algorithm), and FASTA.

Probe sequences found in more than one protein (*e.g.*, A' and B' proteins) are defined as being "linked" so long as at least one protein per domain containing that domain but not the other is also identified. In other words, at least one protein or domain of the plurality of proteins must also be found alone in the sequence database. This verifies that the protein or domain is not an integral part of a first protein but rather a second independent protein having its own functional characteristics.

Statistical methods can be used to judge the significance of possible matches. The statistical significance of an alignment score is described by the probability, P , of obtaining a higher score when the sequences are shuffled. One way to compute a P value threshold is to first consider the total number of sequence comparisons that are to be performed. For example, if there are N proteins in *E. coli* and M in all other genomes this number is $N \times M$. If a comparison of this number of random sequence would result in one pair to yield a P value of $1/NM$ by chance this then is set as the threshold.

This method provides information regarding which proteins are functionally related (*e.g.*, related biological functions common structural complexes, metabolic pathways or biological process) a subset of which physically interact in an organism.

Alignment Algorithms

To align sequences, a number of different procedures can be used that produce a good match between the corresponding residues in the sequences. Typically, the Smith-Waterman (Smith (1981) Adv. Appl. Math. 2:482) or Needleman-Wunsch algorithm (Needleman (1970) J. Mol. Biol. 48:443) algorithm, are used, however, other, faster procedures such as BLAST, FASTA, PSI-BLAST (a version of Blast for finding protein families), or others known in the art (see infra discussion), can be used.

Filtering Methods

The Rosetta Stone Method provides at least two pieces of information. First the method provides information regarding which proteins are functionally related. Second the method provides information regarding which proteins are physically related. Each of these two pieces of information has different sources of error and prediction. The first type of error is introduced by protein sequences that occur in many different proteins and paired with many other protein sequences. The second type of error is introduced due to there often being multiple copies of similar proteins, called paralogs, in a single organism. In general, the "Rosetta Stone" method predicts functionally related proteins well, with no filtering of results required. However, it is possible to filter the error associated with either the first or second type of information.

The invention recognizes that a few domains are linked to an excessive number of other domains by a "Rosetta Stone" protein. For example, 95% of the domains are linked to fewer than 25 other domains. However, some domains, e.g., the Src Homology 3 (SH3) domain or ATP-binding cassette (ABC domains), link to more than a hundred other domains. These links were filtered by removing all links generated involving these 5% of domains (i.e., the domains linked to more than 25 other domains). For example, in *E. coli*, without filtering, 3531 links were identified using the domain-based analysis, but after filtering only 749 links were identified. This method improved prediction of functionally related proteins by 28% and physically related proteins by 47%. Accordingly, there are a number of ways to filter the results to improve the significance of the functional links. As described above, as the number of functional links increases there is an increased higher chance of finding a "Rosetta Stone" protein. By reducing the excessively linked proteins one

reduces the chance number of "Rosetta Stone" proteins thereby increasing the significance of a functional link.

Error introduced by multiple paralogs of linked proteins should have little effect on functional prediction, as paralogs usually have very similar function, but will affect the reliability of prediction of protein-protein interactions. This estimate is calculated for each linked protein pair, and can be estimated roughly as:

$$\text{Fractional Error} = 1 - \frac{\sqrt{N}}{N},$$

where N is the number of paralogous protein pairs, (e.g., A linked to B, A' linked to B', A linked to B', and A' linked to B, in the case that A and A' are paralogs, as are B and B', and the linking proteins is AB as above).

The error can also be estimated as $1-T$, where T is the mean percent of potential true positives calculated for all domain pairs in an organism. For each domain pair linked by a Rosetta Stone protein, there are n proteins with the first domain but not the second, and m proteins with the second domain but not the first. The percent of true positives T is therefore estimated as the smaller of n or m divided by n times m . As this error T can be calculated for each set of linked domains, it can describe the confidence in any particular predicted interaction.

In addition, the error in functional links can be caused by small conserved regions or repeated common amino acid sequences being repeatedly identified in a "Rosetta Stone" protein by a plurality of distinct non-homologous polypeptides. To reduce this error the percent of identity between the "Rosetta Stone" and the distinct non-homologous polypeptide can be measured. Alignment percentages of about 50% to about 90%, or, alternatively, about 75%, between the "Rosetta Stone" and the distinct polypeptide are indicative of links that are not subject to the small peptide sequence.

Phylogenetic Pathway Method

The "phylogenetic profile" method compares protein sequences across all known genomes and analyzes the pattern of inheritance of each protein across the different organisms. In its simplest form, each protein is simply characterized by its presence or absence in each organism. For example, if there are 16 known genomes, then each protein may be assigned a 16-bit code or phylogenetic profile. Since proteins that function together

(e.g., in the same metabolic pathway or as part of a larger functional or structural complex) evolve in a correlated fashion, they should have the same or similar patterns of inheritance, and therefore similar phylogenetic profiles. Therefore, the function of one protein may be inferred from the function of another protein, which has a similar profile, if its function is known. As with the Rosetta Stone method, the function of one protein is inferred from the function of another protein which is dissimilar in sequence. Furthermore, the predicted link between the proteins has utility in developing, for example, drug targets, diagnostics and therapeutics.

The phylogenetic profile method can be implemented in a binary code (*i.e.*, describing the presence or absence of a given protein in an organism) or a continuous code that describes how similar the related sequences are in the different genomes. In addition, grouping of similar protein profiles may be made wherein similar profiles are indicative of functionally related proteins. Furthermore, the requirements for similarity can be modified depending upon particular criteria by varying the difference in similar bit requirements. For example, criteria requiring that the degree of similarity in the profile include all 16 bits being identical can be set, but may be modified so that similarity in 15 bits of the 16 bits would indicate relatedness of the protein profiles as well. Statistical methods can be used to determine how similar two patterns must be in order to be related.

The phylogenetic profile method is applicable to any genome including, e.g., viral, bacterial, archaeal or eukaryotic. The method of phylogenetic profile grouping provides the prediction of function for a previously uncharacterized protein(s). The method also allows prediction of new functional roles for characterized proteins based upon functional linkages. It also provides potential informative connections (*i.e.*, links) between uncharacterized proteins.

To represent the subset of organisms that contain a homolog a phylogenetic profile is constructed for each protein. The simplest manner to represent a protein's phylogenetic history is via a binary phylogenetic profile for each protein. This profile is a string with N entries, each one bit, where N corresponds to the number of genomes. The number of genomes can be any number of two or more (e.g., 2, 3, 4, 5, 10, 100, to 1000 or more). The presence of a homolog to a given protein in the n^{th} genome is indicated with an entry of unity at the n^{th} position (e.g., in a binary system an entry of 1). If no homolog is

found the entry is zero. Proteins are clustered according to the similarity of their phylogenetic profiles. Similar profiles show a correlated pattern of inheritance, and by implication, functional linkage. The method predicts that the functions of uncharacterized proteins are likely to be similar to characterized proteins within a cluster.

5 In order to decide whether a genome contains a protein related to another particular protein, the query amino acid sequence is aligned with each of the proteins from the genome(s) in question using known alignment algorithm (see above). To determine the statistical significance of any alignment score, the probability, p , of obtaining a higher score when the sequences are shuffled is described. One way to compute a p value threshold is to first consider the total number of sequence comparisons that are being aligned. If there are N 10 proteins in a first organism's genome and M in all other genomes this number is $N \times M$. If this number were compared to random sequences it would be expected that one pair would yield a p value of $\frac{1}{NM}$. This value can be set as a threshold. Other thresholds may be used and will be recognized by those of skill in the art.

15 A non-binary phylogenetic profile can be used. In this method, the phylogenetic profile is a string of N entries where the n^{th} entry represents the evolutionary distance of the query protein to the homolog in the n^{th} genome. To define an evolutionary distance between two sequences an alignment between two sequences is performed. Such alignments can be carried out by any number of algorithms known in the art (for examples, 20 see those described above). The evolution is represented by a Markov process with substitution rates, over a fixed interval of time, given by a conditional probability matrix:

$$p(aa \rightarrow aa')$$

where aa and aa' are any amino acids. One way to construct such a matrix is to 25 convert the BLOSUM62 amino acid substitutions matrix (or any other amino acid substitution matrix, e.g., PAM100, PAM250) from a log odds matrix to a conditional probability (or transition) matrix:

$$P_B(i \rightarrow j) = p(j)2 \frac{\text{BLOSUM62}_{ij}}{2} \quad (1)$$

30 $P(i \rightarrow j)$ is the probability that amino acid i will be replaced by amino acid j through point mutations according to the BLOSUM62 scores. The p_j 's are the abundances of amino

acid j and are computed by solving the 20 linear equations given by the normalization conditions that:

$$\sum_i P_B(i \rightarrow j) = 1 \quad (2)$$

5 The probability of this process is computed to account for the observed alignment by taking the product of the conditional probabilities for each aligned pair:

$$P(p) = \prod_n p(aa_n \rightarrow aa'_n) \quad (3)$$

10 A family of evolutionary models is then tested by taking powers of the conditional probability matrix: $p' = p^\alpha(aa \rightarrow aa')$. The power α that maximized P is defined to be the evolutionary distance.

Many other schemes may be imagined to deduce the evolutionary distance between two sequences. For example, one might simply count the number of positions in the sequence where the two proteins have adapted different amino acids.

15 Although the phylogenetic history of an organism can be presented as a vector (as described above), the phylogenetic profiles need not be vectors, but may be represented by matrices. This matrix includes all the pair wise distances between a group of homologous protein, each one from a different organism. Similarly, phylogenetic profiles could be represented as evolutionary trees of homologous proteins. Functional proteins could then be
20 clustered or grouped by matching similar trees, rather than vectors or matrices.

In order to predict function, different proteins are grouped or clustered according to the similarity of their phylogenetic profiles. Similar profiles indicate a correlated pattern of inheritance, and by implication, functional linkage.

25 Grouping or clustering may be accomplished in many ways. The simplest is to compute the Euclidean distance between two profiles. Another method is to compute a correlation coefficient to quantify the similarity between two profiles. All profiles within a specified distance of the query profile are considered to be a cluster or group.

Typically a genome database will be used as a source of sequence information. Where the genome database contains only the nucleic acid sequence that
30 sequence is translated to an amino acid sequence in frame (if known) or in all frames if unknown. Direct comparison of the nucleic acid sequences of two or more organisms may be feasible but will likely be more difficult due to the degeneracy of the genetic code.

Programs capable of translating a nucleic acid sequence are known in the art or easily programmed by those of skill in the art to recognize a codon sequence for each amino acid.

The phylogenetic profile provides an indication of those proteins in each of the at least two organisms that share some degree of homology. Such a comparison can be done by any number of alignment algorithms known in the art or easily developed by one skilled in the art (see, for example, those listed above, *e.g.*, BLAST, FASTA etc.) In addition, thresholds can be set regarding a required degree of homology. Each protein is then grouped at 224 with related proteins that share a similar phylogenetic profile using grouping algorithms.

10 **“Functionally-, Structurally- or Metabolically- Linked” Method**

The “physiologic linkage” method is a computational method that detects (*i.e.*, identifies) proteins, and the genes that encode them, that participate in a common functional pathway (*e.g.*, cell motility or cell division), that participate in the synthesis of the same or a similar structural complex (*e.g.*, a cell wall) or participate in the same or similar metabolic pathway (*e.g.*, glycolysis, lipid synthesis, and the like). Proteins within these common functional pathway groups are examples of “functionally linked” proteins. Having a common functional “goal” they evolve in a correlated fashion. Thus, “homologs” in different organisms can be comparatively identified. While these detection methods are very effective in identifying functional homologues in the same subset of organisms, functional linkages can be made between widely genetically disparate organisms.

In one aspect, metabolic pathways are defined as links between proteins that operate in the same metabolic pathway that can be identified by sequence identity searching, *e.g.*, by performing a BLAST search to find top-scoring polypeptides with high similarity (BLAST alignment E-value $< 10^{-20}$) to polypeptides identified in a known pathway. For example, *M. tuberculosis* proteins were so analyzed against *E. coli* proteins; MTB proteins whose *E. coli* homologs (*i.e.*, having high similarity by BLAST alignment) act adjacently in metabolic pathways as defined in the EcoCyc database (see, *e.g.*, Karp (1998) Nucleic Acids Res. 26:50-53) were identified.

In another example, flagellar proteins are found in bacteria that possess flagella but not in other organisms. Accordingly, if two proteins have homologs in the same subset of fully sequenced organisms, they are likely to be functionally linked. The methods

of the invention use this concept to systematically map links between all the proteins coded by a genome.

Typically, functionally linked proteins have no amino acid sequence similarity with each other and, therefore, cannot be linked by conventional sequence alignment techniques. Accordingly, the methods of the invention identify drug targets that could not be identified using conventional sequence comparison (i.e., sequence homology or sequence identity) techniques.

Prediction of functionally linked proteins by the "phylogenetic method" can also be used in conjunction with the "domain fusion" or "Rosetta Stone" method and also can be filtered by other methods that predict functionally linked proteins, such as the protein phylogenetic profile method or the analysis of correlated mRNA expression patterns. It was found that filtering by these two methods for the Rosetta Stone prediction for *S. cerevisiae*, that proteins predicted to be functionally linked by two or more of these three methods were as likely to be functionally related as proteins that were observed to physically interact by experimental techniques like yeast 2-hybrid methods or co-immunoprecipitation methods.

For example, a combination of these methods of prediction can be used to establish links between proteins of closely related function. The methods of the invention (i.e., the "Rosetta Stone" method and the "phylogenetic profile" method) can be combined with one another or with other protein prediction methods known in the art; see, for example, Eisen (1998) "Cluster analysis and display of genome-wide expression partners," *Proc. Natl. Acad. Sci. USA*, 95:14863-14868.

The various techniques, methods, and variations thereof described can be implemented in part or in whole using computer-based systems and methods. Additionally, computer-based systems and methods can be used to augment or enhance the functionality described above, increase the speed at which the functions can be performed, and provide additional features and aspects as a part of or in addition to those of the invention described elsewhere in this document. Various computer-based systems, methods, and implementations in accordance with this technology are described herein.

Proteins linked to current drug targets

The invention also provides a novel method for identifying a polypeptide, or the nucleic acid sequence that encodes it, that is a target for a drug. The method analyzes the

functional relationship between at least two sequences, wherein at least one of the sequences is a known target of a drug or encodes a polypeptide drug target. The method comprises identifying proteins, and the genes that encode them, that are functionally linked to the targets of known drugs. The functional linkage is determined by using the "domain fusion" method, the "phylogenetic profile" method or the "physiologic linkage" method, or a combination thereof, as described herein.

Thus, this aspect of the invention provides methods identifying drug targets from among all or a subset of genes in a genome using computationally-determined functional linkages. In one implementation of the method, functional linkages are calculated using the "domain fusion" method, the "phylogenetic profile" method or the "physiologic linkage" method, or a combination thereof, between all "query genome genes." Next, each set of genes predicted to be functionally linked to either a known drug target or to a sequence homolog or ortholog (defined below) to a known drug target are examined. These proteins (and the nucleic acids that encode them) are functionally linked to known drug targets; thus, they are operating in the same pathways or systems targeted by the known drug. Accordingly, the methods of the invention have identified them as drug targets.

This method is particularly effective for identifying drug targets in pathogens, such as microorganisms, e.g., bacteria, viruses and the like. This method allows for the identification of novel drug targets that cannot be identified by other techniques, such as traditional sequence homology or sequence identity comparison techniques. Several known drug targets in *M. tuberculosis* were used with the methods of the invention to use functional linkages to identify potential new drug targets in the same pathways as the known drug targets.

There are very few drugs that are effective for anti-tuberculosis therapy, since the complex lipid-rich mycobacterial cell wall is impermeable to many antibacterial agents. Additionally, single- and multi-drug resistance is rapidly emerging against these drugs. To address this issue, the methods of the invention were used to identify *Mycobacterium tuberculosis* (MTB or TB) proteins that are functionally linked to the targets of known drugs. Inhibiting these proteins should have the same effect on the organism as the drug, since the same processes or pathways would be disrupted. Targeting multiple components of a given biochemical pathway would also diminish the opportunity for the development of resistance

because various related proteins would have to mutate against inhibitors while preserving the overall functionality of the pathway.

A list of targets of essential anti-TB drugs (World Health Organization, Geneva, Switzerland) was compiled. The anti-TB drugs included isoniazid, rifampicin, ethambutol, streptomycin, pyrazinamide and thiacetazone. Although not enough is known about the molecular basis of action of the latter two, the functional linkages of the known drug targets was examined.

Isoniazid. This is one of the most widely used of all anti-tuberculosis drugs. It is believed that the compound is activated by the catalase-peroxidase KatG. Once activated, it then attaches to a nicotinamide adenine dinucleotide bound to the enoyl-acyl carrier protein reductase InhA, resulting in the inhibition of mycolic acid biosynthesis Rozwarski (1998) Science 279:98-102.

Using the "phylogenetic profile, the inhA gene was "linked," or functionally associated with, to two polyketide synthases, pks1 and pks6 (Figure 1), both of which contain acyl carrier protein motifs. The polyketide synthase pks6 is in turn known from established metabolic pathways to be linked to fatty acid biosynthesis gene accD3. Further, pks6 is linked to fadD28 and to the operon containing the genes ppsA-E, all recently reported to be crucial for bacterial replication in host lungs (see, e.g., Cox (1999) Nature 402:79-83).

The inhA gene was also linked to an operon encoding two putative oxidoreductases and a gene of entirely unknown function. The inhA gene was further linked to a second operon that includes pepR and gpsI. PepR is a protease whose *Bacillus subtilis* homolog is adjacent to the genes coding for enzymes that synthesize diaminopimelate, a component of the cell wall incorporated by the murE gene product and diaminopicolinate (see, e.g., Chen (1993) J. Biol. Chem. 268:9448-9465). PepR is an ortholog of an essential yeast gene and is likely to be essential for MTB (see below). GpsI is a putative multifunctional enzyme involved in guanosine pentaphosphate synthesis and polyribonucleotide nucleotidyltransfer. The high reliability of the predicted functional link between gpsI and pepR and the absence of eukaryotic homologs suggests that gpsI could be a promising target for drug design.

Rifampicin. This compound, along with the related rifabutin and KRM-1648 are believed to act by directly targeting the RNA polymerase β -subunit (rpoB) given that

96% of resistant isolates were found to have mutations of various types in a limited region of the *rpoB* gene (see, e.g., Yang (1998) J. Antimicrob. Chemother. 42:621-628).

Using the methods of the invention, as expected, functional linkages were found to another RNA polymerase subunit, *rpoC*, as well as to various tRNA synthases and ribosomal proteins. However, no functional links to uncharacterized proteins were found.

Ethambutol. This drug is effective against tuberculosis when used in combination with isoniazid. It is believed that the drug interacts with the *EmbB* protein, a probable arabinosyl-transferase, inhibiting the biosynthesis of arabinan, a component of cell-envelope lipids. As with rifampicin, the evidence for this interaction is indirect, since mutations in the *embB* gene are responsible for ethambutol resistance (see, e.g., Lety (1997) Antimicrob. Agents Chemother. 41:2629-2633).

The "gene proximity" method correctly clusters *embB* with *embA* (Rv3794). This cluster is linked to a set of mostly uncharacterized genes by the "phylogenetic profile" method; see Figure 2, which shows an analysis of *EmbB*, the target for the anti-tuberculosis drug Ethambutol, and shows functional linkages to genes mostly of unknown function but with some indications of localization at the bacterial membrane.

Two of the uncharacterized genes, Rv1706c and Rv1800, belong to the abundant PE/PPE family of proteins hypothesized to be a source of antigenic variation with the potential ability to interfere with immune responses by inhibiting antigen processing (see, e.g., Cole (1998) Nature 393, 537-544). A third uncharacterized gene, Rv1967 belongs to the one of the four copies of the *mce* operon. This operon consists of eight genes coding for integral membrane proteins and proteins that have N-terminal signal sequences or hydrophobic segments and are believed to be involved in pathogenicity (see, e.g., Cole (1998) supra). Rv0528 codes for a hypothetical membrane protein and Rv2159c corresponds to the *murF* gene, which participates in the biosynthesis of peptidoglycan precursors.

The majority of the "links," or functionally associated sequences, involved proteins associated with processes related to the bacterial cell wall (with the possible exception of *atsA* and the putative choline dehydrogenase Rv1279, whose relationship to these processes is not immediately obvious). The proteins of unknown function are therefore also expected to play some role in these processes and are thus of interest as potential drug targets.

Streptomycin. This drug acts by binding to the 16S rRNA and inhibits protein synthesis. Resistance to this compound emerges from mutations in the corresponding gene (*rrs*), as well as in the gene encoding for the ribosomal protein S12 (*rpsL*). Disruptions to *RpsL* effect streptomycin resistance by altering the higher order structure of 16S rRNA (see, e.g., Sreevatsan (1996) *Antimicrob. Agents Chemother.* 40:1024-1026).

Although streptomycin doesn't directly target *RpsL*, the functional links generated for this protein was examined, as any target whose inhibition will ultimately disrupt bacterial protein synthesis is likely to be an effective antigrowth/ anti-microbial target. As with the rifampicin target, the only functional linkages found for this protein were the expected protein synthesis-related proteins, including large ribosomal subunit proteins L2, L5, L11, and L14; small ribosomal subunit proteins S4, S5, S7, S8, and S11; elongation factors *fusA* and *Ef-Tu*; the chaperones *GroEL*, *clpB* and *fisH*; and the *Clp* protease subunits *clpC* and *clpX*.

Proteins linked to cell-wall related proteins

The invention also provides a novel method for identifying a nucleic acid or a polypeptide sequence in an organism that is linked to a cell-wall related protein. The method analyzes the functional relationship between at least two sequences, wherein at least one of the sequences is a cell-wall related protein, or, the sequence is a nucleic acid sequence that encodes a cell-wall related protein. The method comprises identifying proteins, and the genes that encode them, that are functionally linked to a cell-wall related protein. The functional linkage is determined by using the "domain fusion" method, the "phylogenetic profile" method or the "physiologic linkage" method, or a combination thereof, as described herein.

Approximately eleven *M. tuberculosis* proteins are indicated by sequence homology to be penicillin-binding proteins, thought to synthesize peptidoglycan in the course of cell elongation and cell wall metabolism (see, e.g., Broome-Smith (1985) *Eur. J. Biochem.* 147:437-446). Using the methods of the invention, the functional linkages found for these proteins map out many of the known cell wall synthetic enzymes and reveal more than 10 proteins of unknown function that may also participate in cell wall metabolism. Figure 3 shows an analysis of five of the approximately eleven MTB proteins presumed to bind penicillin to reveal functional linkages to various potential operons consisting of genes

involved in various aspects of cell wall metabolism, including cell shape determination and peptidoglycan biosynthesis, as well more than ten genes of unknown function, which we can now associate with cell wall metabolism.

Three of the proteins (pbpA, pbpB, and ponA1) reside in conserved gene clusters, presumably operons. Other genes in the clusters around pbpA and pbpB are also implicated in cell wall metabolism. For example, pbpA resides next to rodA, a membrane-associated protein whose *E. coli* homolog determines cell shape and is required for enzymatic activity of penicillin binding proteins (see, e.g., Matsuzawa (1989) J. Bacteriol. 171:558-560). Likewise, pbpB resides next to six peptidoglycan biosynthesis genes and the two septum and cell wall formation proteins ftsW and ftsZ.

Two additional gene clusters were linked to these penicillin binding proteins by either the "phylogenetic profile" or "Rosetta Stone" pattern methods of the invention. One cluster is composed of the peptidoglycan synthetic protein murB and a putative membrane protein of unknown function that the functional linkages suggest is involved in cell wall metabolism. The second gene cluster contains four genes, three of which are predicted to reside in the cell membrane or envelope. Therefore, the uncharacterized genes in these clusters are likely to be involved in cell wall metabolism, closely related to the function of the penicillin binding proteins and are therefore promising drug targets.

Another gene linked to cell wall metabolism by the computationally-derived linkage methods of the invention is gcpE, see Figure 4, which shows that the uncharacterized gene gcpE, known to be essential for bacterial survival (see, e.g., Baker (1992) FEMS Microbiol. Lett. 73:175-180), is predicted to be involved in cell wall metabolism through its functional links to a putative membrane protein and two murein hydrolase genes, lytB1 and lytB2, involved in cell separation. The genes forming a putative operon with gcpE are proposed as potential drug targets. The functional linkages place gcpE in a conserved gene cluster with two genes of unknown function, one of which encodes a membrane protein. However, the three genes show correlated inheritance with two homologs of lytB, an *E. coli* gene involved in penicillin tolerance (see, e.g., Gustafson (1993) J. Bacteriol. 175:1203-1205) and recently shown to encode a murein hydrolase essential for cell separation (see, e.g., Garcia (1999) Mol. Microbiol. 31:1275-1277). The uncharacterized proteins from this

cluster are therefore expected to participate in processes similar to GcpE and might therefore be promising drug targets.

Proteins linked to potentially novel pathways

The invention also provides a novel method for identifying a polypeptide, or a nucleic acid that encodes it, that is linked to potentially novel biochemical (e.g., biosynthetic, metabolic) pathways. The method analyzes the functional relationship between at least two sequences, wherein at least one of the sequences is associated with a biochemical pathway, such as a pathway in a microorganism that enables the pathogen to evade an immune process. The method comprises identifying proteins, and the genes that encode them, that are functionally linked to the pathway-linked sequences. The functional linkage is determined by using the "domain fusion" method, the "phylogenetic profile" method or the "physiologic linkage" method, or a combination thereof, as described herein.

For example, the *htrA* gene encodes for a putative heat shock protein homologous to HtrA from *Salmonella typhimurium*, a serine protease that degrades aberrant periplasmic proteins. Mutations in this protein have been linked with reduced viability in host macrophages (see, e.g., Johnson (1991) Mol. Microbiol. 5:401-407). Thus, it was decided to investigate the function of *htrA*. Using the methods of the invention, results indicated that the *htrA* protein is part of a process that has not yet been characterized. The gene is predicted with very high reliability to function with the uncharacterized gene Rv1224c, see Figure 5, which shows the involvement of *htrA* in a potentially novel pathway and the gene encoding the putative heat shock protein HtrA is functionally linked to a set of genes mostly of unknown function, suggesting the existence of a novel pathway. The partially characterized proteins suggest that the pathway relates to membrane-associated processes such as signaling and/or transport. The lack of eukaryotic homologs for most of the genes linked to *htrA*, suggests that proteins of this pathway could be promising drug targets.

Through its phylogenetic profile, *htrA* is linked to a group of uncharacterized proteins, including a putative lipid esterase (Rv1900c), an ABC transporter (Rv3783) and the uncharacterized protein Rv1216c, which has weak homology to the laminin B receptor of *Xenopus laevis*, suggesting that it might be a membrane protein. From this analysis, it can be concluded that *htrA* is part of a novel pathway that involves membrane-associated processes,

such as signaling and/or transport. Because the majority of the proteins linked to htrA have no eukaryotic homologs, and given the importance of htrA in *S. typhimurium* pathogenesis, this pathway represents another potential source of novel targets for anti-tuberculosis drugs.

Proteins linked to essential proteins

5 The invention also provides a novel method for identifying a polypeptide, or the nucleic acid sequence that encodes it, that is linked to an essential protein (e.g., a protein necessary for the growth of an organism, such as a bacterium). The method analyzes the functional relationship between at least two sequences, wherein at least one of the sequences is linked to an essential protein, or, the sequence is a nucleic acid sequence that itself is
10 essential or encodes a polypeptide linked to an essential protein. The functional linkage is determined by using the "domain fusion" method, the "phylogenetic profile" method or the "physiologic linkage" method, or a combination thereof, as described herein.

 For example, the MIPS database (Munich Information Center for Protein Sequences; MIPS provides access through its WWW server to a spectrum of generic
15 databases, including PEDANT, MYGD, MATD, MEST, the PIR-International Protein Sequence Database, the protein family database PROTFAM, the MITOP database, and the all-against-all FASTA database; see, e.g., Mewes (1999) Nucleic Acids Res. 27:44-48) contains a list of 734 genes that are essential for *Saccharomyces cerevisiae* viability (see, e.g., Mewes (1999) supra). A list of *Mycobacterium tuberculosis* genes orthologous to these
20 essential genes was generated. Using the methods of the invention, 60 such genes were found. The products of these genes have a high likelihood of also being essential to the tuberculosis bacterium and therefore could be promising therapeutic targets. Furthermore, since the list of essential genes came from a eukaryote, there is a significant chance that these genes would also be found in the human genome.

Automatic Method to Identify Drug Targets from Functional Linkages

 One aspect of the invention provides a computational method to identify potential drug targets among the proteins expressed by a genome. This aspect takes advantage of the functional linkages calculated between genes in a genome using the methods described herein, as well as the detection of sequence homology and the knowledge
30 of a set of lethal or "essential" genes in one or more organisms.

To identify drug targets in a query genome, the sequence homology between all of the genes in that genome and all of the genes in the genome of an organism for which essential genes are known is calculated. For example, as discussed herein, the query genome is *Mycobacterium tuberculosis* (TB) and the genome with known essentials is the yeast *S. cerevisiae*. Sequence homology between all TB genes and all yeast genes was calculated using the methods of the invention.

"Equivalent" or "orthologous" genes were also identified by another aspect of the invention that comprises doing a reverse sequence search (e.g., yeast vs. TB) and then choosing pairs of genes that are the symmetric best-scoring sequence search. In one exemplary aspect, MTB orthologs of *Saccharomyces cerevisiae* genes were generated by finding all pairs of genes (TB_i, SC_j) where TB_i was the top hit from a BLAST search of the yeast gene SC_j against the MTB genome, SC_j was the top hit from a BLAST search of the MTB gene TB_i against the *Saccharomyces cerevisiae* genome and both top hits had a BLAST E-value $\leq 1 \times 10^{-5}$.

For example, a TB gene is an ortholog of a yeast gene if the yeast gene is the best scoring sequence match when yeast is searched with the TB gene, and the TB gene is the best scoring sequence match when TB is searched with the yeast gene. We define these "symmetric" pairs as "orthologs."

After identifying orthologs between the query genome and the genome with known essential genes, a set of query genome genes that are orthologs of known essential genes in the other genome was chosen. These genes were designated the set of "putative essentials". For the purposes of the algorithm of the invention, these query genome genes are assumed to be essential genes, since they are the equivalents of essential genes in another genome. These genes act as "markers" or indicators of essential pathways in the query genome. One could supplement this set with genes already known to be essential in the query organism. Functional linkages (determined by the methods of the invention) between all query genome genes were examined. The query genome genes linked to all of the putative essential genes were examined. This set of genes was designated as the "predicted members of essential pathways." These genes are likely to be involved in important pathways, since the (predicted) pathways have members that are putative essentials. Lastly, the method removes from the set of genes in predicted essential pathways all of those genes

that have sequence homology to eukaryotic genes or proteins. The genes that remain after this filtering step are the predicted drug targets for the query organism.

As a benchmark, this method was applied to the *M. tuberculosis* genome. Of the over 3900 genes in TB, 11 were identified as potential drug targets. Comparing this list of 11 predicted targets to the less than 10 known drug anti-TB drug targets, one gene was a known drug target and one was linked to a known drug target. Accordingly, the algorithm of the invention performed statistically significantly much better than a random choice of genes. A rough estimate of statistical significance suggests that one would expect to see 2 of 10 known drug targets in a sample of 11 out of 3900 genes only 3.8 times out of 10,000 trials (probability of occurring by random chance of 3.8×10^{-4}). Therefore, this embodiment of the method is an entirely computational algorithm drawing on the demonstrated ability of the general methods of the invention to predict functional linkages between genes and to effectively identify drug targets in bacteria. The effectiveness of this method to identify novel drug targets was clearly demonstrated when the algorithm was applied to the *M. tuberculosis* genome.

The specific inhibition of the MTB homologs might be difficult. To address this issue, using the methods of the invention, functional links to the essential genes were searched. Functional links were selected which either do not have homologs in yeast, or the enzymatic activity of their products are known to be absent in human cells. Using the highest confidence data, functional links for 23 of the genes (indicated in **bold** in Table 1) were found.

Table 1. MTB orthologs of essential yeast proteins.

Name [†]	Gene	Comments	Name	Gene	Comments
Rv0005 [†]	gyrB	DNA gyrase subunit B	Rv2101	hel2	probable helicase, Sm2/Rad54 family
Rv0014c	pkxB	serine-threonine protein kinase	Rv2110c	prcB	proteasome [beta]-type subunit 2
Rv0032	bioF2	C-terminal similar to B. subtilis BioF	Rv2118c	-	= B2126_C1_165 (83.6%)
Rv0350	dnk	70 kD heat shock protein, chromosome replication	Rv2438c	-	similar to YHNA_YEAST P38795
Rv0363c	fba	fructose biphosphate aldolase	Rv2439c	prcB	glutamate 5-kinase
Rv0435c	-	ATPase of AAA-family	Rv2448c	valS	valyl-tRNA synthase
Rv0436c	psaA	CDP-diacylglycerol-serine o-phosphatidyltransferase	Rv2509	-	putative oxidoreductase
Rv0440	groEL2	60 kD chaperonin 2	Rv2524c	fes	fatty acid synthase
Rv0489	gpm	phosphoglycerate mutase I	Rv2555c	als	alanyl-tRNA synthase
Rv0490	senX3	sensor histidine kinase	Rv2580c	hisS	histidyl-tRNA synthase
Rv0500	proC	pyrroline-5-carboxylate reductase	Rv2614c	thrS	threonyl-tRNA synthase
Rv0667	rpoB	[beta] subunit of RNA polymerase	Rv2697c	dut	deoxyuridine triphosphatase
Rv0668	rpoC	[beta] subunit of RNA polymerase	Rv2782c	pepR	protease/peptidase, M16 family (insulinase)
Rv0764c	-	possible lanosterol 14-demethylase cytochrome P450	Rv2793c	trub	tRNA pseudouridine 55 synthase
Rv0861c	-	probable DNA helicase	Rv2922c	smc	member of Smc1/Cut3/Cut14 family
Rv1010	ksgA	16S rRNA dimethyltransferase	Rv2925c	rnc	RNAse III
Rv1106c	-	probable cholesterol dehydrogenase	Rv3014c	ligA	DNA ligase
Rv1229c	mtp	similar to MRP/NBP35 ATP-binding proteins	Rv3025c	-	Nifs-like protein
Rv1239c	corA	probable magnesium and cobalt transport protein	Rv3080c	pknK	serine-threonine protein kinase
Rv1294	thrA	homoserine dehydrogenase	Rv3106	fprA	adenodoxin and NADPH ferredoxin reductase
Rv1323	fadA4	acetyl-CoA C-acetyltransferase (aka thl)	Rv3255c	manA	mannose-6-phosphate isomerase
Rv1389	gmk	putative guanylate kinase	Rv3264c	rmlA2	glucose-1-phosphate thymidyltransferase
Rv1407	fmu	similar to Fmu protein	Rv3418c	groES	10 kD chaperone
Rv1409	ribG	riboflavin biosynthesis	Rv3490	otsA	probable [alpha]-D-trehalose-phosphate synthase
Rv1617	pykA	pyruvate kinase	Rv3598c	lysS	lysyl-tRNA synthase
Rv1630	rpsA	30S ribosomal protein S1	Rv3608c	folP	dihydropterolate synthase
Rv1745c	-	similar to Q46822 ORF_0182	Rv3609c	folE	GTP cyclohydrolase I
Rv1844c	gnd	6-phosphogluconate dehydrogenase (Gram -)	Rv3721c	dnaZx	DNA polymerase III, [gamma] (dnaZ) and (dnaX)
Rv1981c	ncdF	ribonucleotide reductase small subunit	Rv3834c	serS	seryl-tRNA synthase
Rv2092c	hly	probable helicase, Sk12 subfamily	Rv3907c	pcnA	polynucleotide polymerase

[†] We follow the Sanger Centre naming convention for MTB genes.

* Genes for which high-confidence functional links were found shown in boldface

Eight of these were linked to 12 unique MTB genes that satisfied the criteria of the invention's methods (Table 1). Exemplary findings include:

(1) the gene folP, which encodes the enzyme dihydropteroate synthase (DHPS) known to be the target of sulfonamide antibacterial drugs. Although it is found in some eukaryotes, DHPS activity is not found in human cells (see, e.g., Huovinen (1995) Antimicrob. Agents Chemother. 39:279-2890.

(2) the product of the gene folK, a 7,8-dihydro-6-hydroxymethyl-pterinpyrophosphokinase, has recently been proposed as a target for broad-spectrum antibacterial drugs (see, e.g., Stammers (1999) FEBS Lett. 456:49-53).

(3) the gene gpsI, is not only strongly linked to the essential yeast gene pepR, but it is also functionally linked to inhA, the target of the drug isoniazid (see above), making it a very compelling candidate for drug design.

Table 2. Subset of genes from Table 1 that are functionally linked to genes without yeast homologs.

Gene	Link [†]	Comments
Rv0005	Rv0002	dnaN DNA polymerase III, β -subunit
	Rv0003	recF DNA replication and SOS induction
	Rv0006	gyrA DNA gyrase subunit A
Rv0350	Rv0351	grpE stimulates DnaK ATPase activity
	Rv0352	dnaJ acts with GrpE to stimulate DnaK ATPase
Rv1010	Rv1008	Similar to <i>E.coli</i> hypothetical protein YcfH
	Rv1009	Possible lipoprotein, similar to various other MTB proteins
	Rv1011	Similar to <i>E.coli</i> hypothetical protein YcbH
Rv2439c	Rv2427c	proA γ -glutamyl phosphate reductase
	Rv2440c	obg Obg GTP-binding protein
	Rv2441c	rpmA 50S ribosomal protein L27
	Rv2442c	rplU 50S ribosomal protein L21
Rv2782c	Rv2783c	gpsI pppGpp synthase and polyribonucleotide phosphorylase
Rv3598c	Rv3600c	similar to <i>Bacillus subtilis</i> hypothetical protein YacB
	Rv3606c	folK 7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase
	Rv3607c	folX may be involved in folate biosynthesis
	Rv3608c[‡]	folP dihydropteroate synthase (DHPS)
	Rv3610c	ftsH inner membrane protein, chaperone
Rv3608c	Rv3598c	lysS lysyl-tRNA synthase
	Rv3600c	similar to <i>Bacillus subtilis</i> hypothetical protein YacB
	Rv3606c	folK 7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase
	Rv3607c	folX may be involved in folate biosynthesis
	Rv3609c	folE GTP cyclohydrolase I
	Rv3610c	ftsH inner membrane protein, chaperone
Rv3609c	Rv3606c	folK 7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase
	Rv3607c	folX may be involved in folate biosynthesis
	Rv3608c[‡]	folP dihydropteroate synthase (DHPS)

[†] Genes without yeast homologs shown in boldface[‡] DHPS activity is found in some eukaryotic cells but not in human cells

5 In summary, the methods of the invention allowed identification of this combination of functional linkages to essential genes. This information, together with the lack of eukaryotic homologs for these genes, makes this group of proteins promising drug targets, particularly because their inhibition is expected to disrupt vital bacterial processes with a low likelihood of toxicity from the inhibition of a host equivalent.

Computer Implementation

The various techniques, methods, and aspects of the invention described herein can be implemented in part or in whole using computer-based systems and methods. Additionally, computer-based systems and methods can be used to augment or enhance the functionalities and algorithms described herein, increase the speed at which the functions can be performed, and provide additional features and aspects as a part of or in addition to those of the invention described elsewhere in this document. Various exemplary computer-based systems, methods and implementations in accordance with the above-described technology are presented herein.

The processor-based system can include a main memory, such as a random access memory (RAM), and can also include a secondary memory. The secondary memory can include, for example, a hard disk drive and/or a removable storage drive, representing a floppy disk drive, a magnetic tape drive, an optical disk drive, *etc.* The removable storage drive reads from and/or writes to a removable storage medium. Removable storage media can be a floppy disk magnetic tape, an optical disk, and the like, which can be read by and written to by removable storage drive. The removable storage media can includes a computer usable storage medium having stored therein computer software and/or data.

In alternative embodiments, secondary memory may include other similar means for allowing computer programs or other instructions to be loaded into a computer system. Such means can include, for example, a removable storage unit and an interface. Examples of such can include a program cartridge and cartridge interface (such as the found in video game devices), a movable memory chip (such as an EPROM, or PROM) and associated socket, and other removable storage units and interfaces that allow software and data to be transferred from the removable storage unit to the computer system.

The computer system can also include a communications interface. Communications interfaces allow software and data to be transferred between computer system and external devices. Examples of communications interfaces include modems, network interfaces (such as, for example, an Ethernet card), communications ports, PCMCIA slots and cards, and the like. Software and data transferred via a communications interface can be in the form of signals that can be electronic, electromagnetic, optical or other signals capable of being received by a communications interface. These signals can be provided to

communications interface via a channel capable of carrying signals and can be implemented using a wireless medium, wire or cable, fiber optics or other communications medium. Some examples of a channel can include a phone line, a cellular phone link, an RF link, a network interface, and other communications channels.

5 As used herein, the terms "computer program medium" and "computer usable medium" are used to generally refer to media such as a removable storage device, a disk capable of installation in a disk drive, and signals on a channel, or equivalents thereof. These computer program products are means for providing software or program instructions to computer systems. Computer programs (also called computer control logic) can be stored in
10 main memory and/or secondary memory. Computer programs can also be received via a communications interface. Such computer programs, when executed, enable the computer system to perform the features of the present invention as discussed herein. Computer programs, when executed, enable the processor to perform the features of the present invention. Accordingly, in one aspect of the invention, such computer programs represent
15 controllers of the computer system.

 In another aspect of the invention the methods and algorithms are implemented using software, the software may be stored in, or transmitted via, a computer program product and loaded into a computer system using a removable storage drive, hard drive or communications interface. The control logic (software), when executed by the
20 processor, causes the processor to perform the functions of the invention as described herein.

 In another aspect, the elements are implemented primarily in hardware using, for example, hardware components such as PALs, application specific integrated circuits (ASICs) or other hardware components. Implementation of a hardware state machine so as to perform the functions described herein will be apparent to person skilled in the relevant
25 art(s). In yet another embodiment, elements are implanted using a combination of both hardware and software.

 In another aspect, the computer-based methods can be accessed or implemented over the World Wide Web by providing access via a Web Page to the methods of the present invention. Accordingly, the Web Page is identified by a Universal Resource
30 Locator (URL). The URL denotes both the server machine, and the particular file or page on that machine. In this embodiment, it is envisioned that a consumer or client computer system

interacts with a browser to select a particular URL, which in turn causes the browser to send a request for that URL or page to the server identified in the URL. Typically the server responds to the request by retrieving the requested page, and transmitting the data for that page back to the requesting client computer system (the client/server interaction is typically performed in accordance with the hypertext transport protocol ("HTTP")). The selected page is then displayed to the user on the client's display screen. The client may then cause the server containing a computer program of the present invention to launch an application comprising a method of the invention, for example, to identify a nucleic acid or a polypeptide sequence that may be a target for a drug comprising the steps of (a) providing a first nucleic acid or a polypeptide sequence that is known to be a drug target; (b) providing an algorithm capable analyzing a functional relationship between nucleic acid or polypeptide sequences selected from the group consisting of a "domain fusion" method, a "phylogenetic profile" method and a "physiologic linkage" method; and, (c) comparing the first nucleic acid or the polypeptide drug target sequence to a plurality of sequences using at least one algorithm to identify a second sequence that has a functional relationship to the first sequence, thereby identifying a nucleic acid or a polypeptide sequence that may be a target for a drug, based on a query sequence provided by the client.

Nucleic Acids and Polypeptides

The invention also provides isolated nucleic acids and polypeptides comprising the sequences as set forth in Table 3 and Table 4 (below). As used herein, "isolated," when referring to a molecule or composition, such as, *e.g.*, an isolated infected cell comprising a nucleic acid sequence derived from a library of the invention, means that the molecule or composition (including, *e.g.*, a cell) is separated from at least one other compound, such as a protein, DNA, RNA, or other contaminants with which it is associated *in vivo* or in its naturally occurring state. Thus, a nucleic acid or polypeptide or peptide sequence is considered isolated when it has been isolated from any other component with which it is naturally associated. An isolated composition can, however, also be substantially pure. An isolated composition can be in a homogeneous state. It can be in a dry or an aqueous solution. Purity and homogeneity can be determined, *e.g.*, using any analytical chemistry technique, as described herein.

The term "nucleic acid" or "nucleic acid sequence" refers to a deoxy-ribonucleotide or ribonucleotide oligonucleotide, including single- or double-stranded, or coding or non-coding (e.g., "antisense") forms. The term encompasses nucleic acids, i.e., oligonucleotides, containing known analogues of natural nucleotides. The term also
5 encompasses nucleic-acid-like structures with synthetic backbones, see e.g., *Oligonucleotides and Analogues, a Practical Approach*, ed. F. Eckstein, Oxford Univ. Press (1991); *Antisense Strategies*, Annals of the N.Y. Academy of Sciences, Vol 600, Eds. Baserga et al. (NYAS 1992); Milligan (1993) *J. Med. Chem.* 36:1923-1937; *Antisense Research and Applications* (1993, CRC Press), WO 97/03211; WO 96/39154; Mata (1997) *Toxicol. Appl. Pharmacol.*
10 144:189-197; Strauss-Soukup (1997) *Biochemistry* 36:8692-8698; Samstag (1996) *Antisense Nucleic Acid Drug Dev* 6:153-156. As used herein, the "sequence" of a nucleic acid or gene refers to the order of nucleotides in the polynucleotide, including either or both strands (sense and antisense) of a double-stranded DNA molecule, e.g., the sequence of both the coding strand and its complement, or of a single-stranded nucleic acid molecule (sense or antisense).
15 For example, in alternative embodiments, promoters drive the transcription of sense and/or antisense polynucleotide sequences of the invention, as exemplified by Table 3.

The terms "polypeptide," "protein," and "peptide" include compositions of the invention that also include "analogs," or "conservative variants" and "mimetics" ("peptidomimetics") with structures and activity that substantially correspond to the
20 exemplary sequences, such as the sequences in Table 4. Thus, the terms "conservative variant" or "analog" or "mimetic" also refer to a polypeptide or peptide which has a modified amino acid sequence, such that the change(s) do not substantially alter the polypeptide's (the conservative variant's) structure and/or activity (e.g., immunogenicity, ability to bind to human antibodies, etc.), as defined herein. These include conservatively modified variations
25 of an amino acid sequence, i.e., amino acid substitutions, additions or deletions of those residues that are not critical for protein activity, or substitution of amino acids with residues having similar properties (e.g., acidic, basic, positively or negatively charged, polar or non-polar, etc.) such that the substitutions of even critical amino acids does not substantially alter structure and/or activity. Conservative substitution tables providing functionally similar
30 amino acids are well known in the art. For example, one exemplary guideline to select conservative substitutions includes (original residue followed by exemplary substitution):

ala/gly or ser; arg/ lys; asn/ gln or his; asp/glu; cys/ser; gln/asn; gly/asp; gly/ala or pro; his/asn or gln; ile/leu or val; leu/ile or val; lys/arg or gln or glu; met/leu or tyr or ile; phe/met or leu or tyr; ser/thr; thr/ser; trp/tyr; tyr/trp or phe; val/ile or leu. An alternative exemplary guideline uses the following six groups, each containing amino acids that are conservative substitutions for one another: 1) Alanine (A), Serine (S), Threonine (T); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); (see also, e.g., Creighton (1984) *Proteins*, W.H. Freeman and Company; Schulz and Schirmer (1979) *Principles of Protein Structure*, Springer-Verlag). One of skill in the art will appreciate that the above-identified substitutions are not the only possible conservative substitutions. For example, for some purposes, one may regard all charged amino acids as conservative substitutions for each other whether they are positive or negative. In addition, individual substitutions, deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence can also be considered "conservatively modified variations."

The terms "mimetic" and "peptidomimetic" refer to a synthetic chemical compound that has substantially the same structural and/or functional characteristics of the polypeptides of the invention (e.g., ability to bind, or "capture," human antibodies in an ELISA). The mimetic can be either entirely composed of synthetic, non-natural analogues of amino acids, or, is a chimeric molecule of partly natural peptide amino acids and partly non-natural analogs of amino acids. The mimetic can also incorporate any amount of natural amino acid conservative substitutions as long as such substitutions also do not substantially alter the mimetics' structure and/or activity. As with polypeptides of the invention which are conservative variants, routine experimentation will determine whether a mimetic is within the scope of the invention, i.e., that its structure and/or function is not substantially altered. Polypeptide mimetic compositions can contain any combination of non-natural structural components, which are typically from three structural groups: a) residue linkage groups other than the natural amide bond ("peptide bond") linkages; b) non-natural residues in place of naturally occurring amino acid residues; or c) residues which induce secondary structural mimicry, i.e., to induce or stabilize a secondary structure, e.g., a beta turn, gamma turn, beta sheet, alpha helix conformation, and the like. A polypeptide can be characterized as a

mimetic when all or some of its residues are joined by chemical means other than natural peptide bonds. Individual peptidomimetic residues can be joined by peptide bonds, other chemical bonds or coupling means, such as, e.g., glutaraldehyde, N-hydroxysuccinimide esters, bifunctional maleimides, N,N'-dicyclohexylcarbodiimide (DCC) or N,N'-

5 diisopropylcarbodiimide (DIC). Linking groups that can be an alternative to the traditional amide bond ("peptide bond") linkages include, e.g., ketomethylene (e.g., -C(=O)-CH₂- for -C(=O)-NH-), aminomethylene (CH₂-NH), ethylene, olefin (CH=CH), ether (CH₂-O), thioether (CH₂-S), tetrazole (CN₄-), thiazole, retroamide, thioamide, or ester (see, e.g., Spatola (1983) in Chemistry and Biochemistry of Amino Acids, Peptides and Proteins, Vol.
10 7, pp 267-357, "Peptide Backbone Modifications," Marcell Dekker, NY). A polypeptide can also be characterized as a mimetic by containing all or some non-natural residues in place of naturally occurring amino acid residues; non-natural residues are well described in the scientific and patent literature.

The invention comprises nucleic acids comprising sequences as set forth in
15 Table 3, or comprising nucleic acids encoding the polypeptides as set forth in Table 4, operably linked to a transcriptional regulatory sequence. As used herein, the term "operably linked," refers to a functional relationship between two or more nucleic acid (e.g., DNA) segments. Typically, it refers to the functional relationship of a transcriptional regulatory sequence to a transcribed sequence. For example, a promoter (defined below) is operably
20 linked to a coding sequence, such as a nucleic acid of the invention, if it stimulates or modulates the transcription of the coding sequence in an appropriate host cell or other expression system. Generally, promoter transcriptional regulatory sequences that are operably linked to a transcribed sequence are physically contiguous to the transcribed sequence, *i.e.*, they are *cis*-acting. However, some transcriptional regulatory sequences, such
25 as enhancers, need not be physically contiguous or located in close proximity to the coding sequences whose transcription they enhance. For example, in one embodiment, a promoter is operably linked to an ORF-containing nucleic acid sequence of the invention, as exemplified by, e.g., a nucleic acid sequence as set forth in Table 3.

As used herein, the term "promoter" includes all sequences capable of driving
30 transcription of a coding sequence in an expression system. Thus, promoters used in the constructs of the invention include *cis*-acting transcriptional control elements and regulatory

sequences that are involved in regulating or modulating the timing and/or rate of transcription of a nucleic acid of the invention. For example, a promoter can be a *cis*-acting transcriptional control element, including an enhancer, a promoter, a transcription terminator, an origin of replication, a chromosomal integration sequence, 5' and 3' untranslated regions, or an intronic sequence, which are involved in transcriptional regulation. These *cis*-acting sequences typically interact with proteins or other biomolecules to carry out (turn on/off, regulate, modulate, etc.) transcription.

The invention comprises expression cassettes comprising nucleic acids comprising sequences as set forth in Table 3, or comprising nucleic acids encoding the polypeptides as set forth in Table 4. The term "expression vector" refers to any recombinant expression system for the purpose of expressing a nucleic acid sequence of the invention *in vitro* or *in vivo*, constitutively or inducibly, in any cell, including prokaryotic, yeast, fungal, plant, insect or mammalian cell. The term includes linear or circular expression systems. The term includes expression systems that remain episomal or integrate into the host cell genome. The expression systems can have the ability to self-replicate or not, *i.e.*, drive only transient expression in a cell. The term includes recombinant "expression cassettes" which contain only the minimum elements needed for transcription of the recombinant nucleic acid.

Alignment Analysis of Sequences

The nucleic acid and polypeptide sequences of the invention include genes and gene products identified and characterized by sequence identify analysis (*i.e.*, by homology) using the exemplary nucleic acid and protein sequences of the invention, including, *e.g.*, those set forth in Tables 3 and 4. In alternative aspects of the invention, nucleic acids and polypeptides within the scope of the invention include those having 98%, 95%, 90%, 85% or 80% sequence identity (homology) to the exemplary sequences as set forth in Tables 3 and 4.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters are used unless alternative parameters are designated herein. The sequence comparison algorithm then calculates the percent sequence identity for the test

sequence(s) relative to the reference sequence, based on the designated or default program parameters. A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 25 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence
5 may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970),
10 by the search for similarity method of Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (CLUSTAL, GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection.

15 In one aspect of the invention (in the methods of the invention, and, to determine if a sequence is within the scope of the invention), a CLUSTAL algorithm is used, *e.g.*, the CLUSTAL W program, see, *e.g.*, Thompson (1994) *Nuc. Acids Res.* 22:4673-4680; Higgins (1996) *Methods Enzymol* 266:383-402. Variations can also be used, such as CLUSTAL X, see Jeanmougin (1998) *Trends Biochem Sci* 23:403-405; Thompson (1997)
20 *Nucleic Acids Res* 25:4876-4882. In one aspect, the CLUSTAL W program described by Thompson (1994) *supra*, is used with the following parameters: K tuple (word) size: 1, window size: 5, scoring method: percentage, number of top diagonals: 5, gap penalty: 3, to determine whether a nucleic acid has sufficient sequence identity to an exemplary sequence to be with the scope of the invention. In another aspect, the algorithm PILEUP is used in the
25 methods and to determine whether a nucleic acid has sufficient sequence identity to be with the scope of the invention. This program creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendrogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method
30 of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). Using PILEUP, a

reference sequence (e.g., an exemplary GCA-associated sequence of the invention) is compared to another sequence to determine the percent sequence identity relationship (i.e., that the second sequence is substantially identical and within the scope of the invention) using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. In one embodiment, PILEUP obtained from the GCG sequence analysis software package, e.g., version 7.0 (Devereaux(1984) *Nuc. Acids Res.* 12:387-395), using the parameters described therein, is used in the methods and to identify nucleic acids within the scope of the invention. In a another aspect, a BLAST algorithm is used (in the methods, e.g., to determine percent sequence identity (i.e., substantial similarity or identity) and whether a nucleic acid is within the scope of the invention), see, e.g., Altschul (1990) *J. Mol. Biol.* 215:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information, NIH. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul (1990) *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues, always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. In one embodiment, to determine if a nucleic acid sequence is within the scope of the invention, the BLASTN program (for nucleotide sequences) is used incorporating as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as default parameters a wordlength (W) of 3, an expectation (E) of 10, and the

BLOSUM62 scoring matrix (see, e.g., Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

Hybridization for Identifying Nucleic Acids of the Invention

Nucleic acids within the scope of the invention include isolated or
5 recombinant nucleic acids that specifically hybridize under stringent hybridization conditions to an exemplary nucleic acid of the invention (including a sequence encoding an exemplary polypeptide) as set forth in Tables 3 and 4. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at
10 higher temperatures. An extensive guide to the hybridization of nucleic acids is found in, e.g., Tijssen (1993) *infra*. Generally, stringent conditions are selected to be about 5 to 10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of
15 the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of
20 destabilizing agents such as formamide.

For selective or specific hybridization, a positive signal (e.g., identification of a nucleic acid of the invention) is about 10 times background hybridization. "Stringent" hybridization conditions that are used to identify substantially identical nucleic acids within the scope of the invention include hybridization in a buffer comprising 50% formamide, 5x
25 SSC, and 1% SDS at 42°C, or hybridization in a buffer comprising 5x SSC and 1% SDS at 65°C, both with a wash of 0.2x SSC and 0.1% SDS at 65°C. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, and 1% SDS at 37°C, and a wash in 1X SSC at 45°C. Those of ordinary skill will readily recognize that alternative but comparable hybridization and wash conditions can be
30 utilized to provide conditions of similar stringency. Nucleic acids which do not hybridize to each other under stringent hybridization conditions are still substantially identical if the

polypeptides which they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code, as discussed herein (see discussion on "conservative substitutions"). However, the selection of a hybridization format is not critical - it is the stringency of the wash

5 conditions that set forth the conditions that determine whether a nucleic acid is within the scope of the invention. Wash conditions used to identify nucleic acids within the scope of the invention include, e.g.: a salt concentration of about 0.02 molar at pH 7 and a temperature of at least about 50°C or about 55°C to about 60°C; or, a salt concentration of about 0.15 M NaCl at 72°C for about 15 minutes; or, a salt concentration of about 0.2X SSC at a

10 temperature of at least about 50°C or about 55°C to about 60°C for about 15 to about 20 minutes; or, the hybridization complex is washed twice with a solution with a salt concentration of about 2X SSC containing 0.1% SDS at room temperature for 15 minutes and then washed twice by 0.1X SSC containing 0.1% SDS at 68°C for 15 minutes; or, equivalent conditions. See Sambrook, Tijssen and Ausubel (see below) for a description of

15 SSC buffer and equivalent conditions.

General Techniques

The nucleic acid and polypeptide sequences of the invention and other nucleic acids used to practice this invention, whether RNA, cDNA, genomic DNA, vectors, viruses or hybrids thereof, may be isolated from a variety of sources, genetically engineered,

20 amplified, and/or expressed recombinantly. Any recombinant expression system can be used, including, in addition to bacterial cells, e.g., mammalian, yeast, insect or plant cell expression systems.

Alternatively, these nucleic acids and polypeptides can be synthesized *in vitro* by well-known chemical synthesis techniques, as described in, e.g., Carruthers (1982) Cold

25 Spring Harbor Symp. Quant. Biol. 47:411-418; Adams (1983) J. Am. Chem. Soc. 105:661; Belousov (1997) Nucleic Acids Res. 25:3440-3444; Frenkel (1995) Free Radic. Biol. Med. 19:373-380; Blommers (1994) Biochemistry 33:7886-7896; Narang (1979) Meth. Enzymol. 68:90; Brown (1979) Meth. Enzymol. 68:109; Beaucage (1981) Tetra. Lett. 22:1859; U.S. Patent No. 4,458,066.

30 Techniques for the manipulation of nucleic acids, such as, e.g., generating mutations in sequences, subcloning, labeling probes, sequencing, hybridization and the like

are well described in the scientific and patent literature, see, e.g., Sambrook, ed.,
MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor
Laboratory, (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley
& Sons, Inc., New York (1997); LABORATORY TECHNIQUES IN BIOCHEMISTRY AND
5 MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID PROBES, Part I. Theory and
Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y. (1993).

Polypeptides and peptides of the invention can also be synthesized, whole or
in part, using chemical methods well known in the art. See e.g., Caruthers (1980) Nucleic
Acids Res. Symp. Ser. 215-223; Horn (1980) Nucleic Acids Res. Symp. Ser. 225-232;
10 Banga, A.K., Therapeutic Peptides and Proteins, Formulation, Processing and Delivery
Systems (1995) Technomic Publishing Co., Lancaster, PA. For example, peptide synthesis
can be performed using various solid-phase techniques (see e.g., Roberge (1995) Science
269:202; Merrifield (1997) Methods Enzymol. 289:3-13) and automated synthesis may be
achieved, e.g., using the ABI 431A Peptide Synthesizer (Perkin Elmer) in accordance with
15 the instructions provided by the manufacturer.

The skilled artisan will recognize that individual synthetic residues and
polypeptides incorporating mimetics can be synthesized using a variety of procedures and
methodologies, which are well described in the scientific and patent literature, e.g., Organic
Syntheses Collective Volumes, Gilman, et al. (Eds) John Wiley & Sons, Inc., NY.

20 Polypeptides incorporating mimetics can also be made using solid phase synthetic
procedures, as described, e.g., by Di Marchi, et al., U.S. Pat. No. 5,422,426. Peptides and
peptide mimetics of the invention can also be synthesized using combinatorial
methodologies. Various techniques for generation of peptide and peptidomimetic libraries
are well known, and include, e.g., multipin, tea bag, and split-couple-mix techniques; see,
25 e.g., al-Obeidi (1998) Mol. Biotechnol. 9:205-223; Hruby (1997) Curr. Opin. Chem. Biol.
1:114-119; Ostergaard (1997) Mol. Divers. 3:17-27; Ostresh (1996) Methods Enzymol.
267:220-234. Modified peptides of the invention can be further produced by chemical
modification methods, see, e.g., Belousov (1997) Nucleic Acids Res. 25:3440-3444; Frenkel
(1995) Free Radic. Biol. Med. 19:373-380; Blommers (1994) Biochemistry 33:7886-7896.

30 Peptides and polypeptides of the invention can also be synthesized and
expressed as fusion proteins with one or more additional domains linked thereto for, e.g.,

producing a more immunogenic peptide, to more readily isolate a recombinantly synthesized peptide, to identify and isolate antibodies and antibody-expressing B cells, and the like.

Detection and purification facilitating domains include, *e.g.*, metal chelating peptides such as polyhistidine tracts and histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA). The inclusion of a cleavable linker sequences such as Factor Xa or enterokinase (Invitrogen, San Diego CA) between the purification domain and GCA-associated peptide or polypeptide can be useful to facilitate purification. For example, an expression vector can include an epitope-encoding nucleic acid sequence linked to six histidine residues followed by a thioredoxin and an enterokinase cleavage site (see *e.g.*, Williams (1995) *Biochemistry* 34:1787-1797; Dobeli (1998) *Protein Expr. Purif.* 12:404-414). The histidine residues facilitate detection and purification while the enterokinase cleavage site provides a means for purifying the epitope from the remainder of the fusion protein. Technology pertaining to vectors encoding fusion proteins and application of fusion proteins are well described in the scientific and patent literature, see *e.g.*, Kroll (1993) *DNA Cell. Biol.*, 12:441-53.

The invention provides antibodies that specifically bind to the polypeptides of the invention, as set forth in Table 4. These antibodies can be useful in the screening methods of the invention. The polypeptides or peptide can be conjugated to another molecule or can be administered with an adjuvant. The coding sequence can be part of an expression cassette or vector capable of expressing the immunogen *in vivo*. (see, *e.g.*, Katsumi (1994) *Hum. Gene Ther.* 5:1335-9). Methods of producing polyclonal and monoclonal antibodies are known to those of skill in the art and described in the scientific and patent literature, see, *e.g.*, Coligan, *CURRENT PROTOCOLS IN IMMUNOLOGY*, Wiley/Greene, NY (1991); Stites (eds.) *BASIC AND CLINICAL IMMUNOLOGY* (7th ed.) Lange Medical Publications, Los Altos, CA; Goding, *MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE* (2d ed.) Academic Press, New York, NY (1986); Harlow (1988) *ANTIBODIES, A LABORATORY MANUAL*, Cold Spring Harbor Publications, New York.

Antibodies also can be generated *in vitro*, *e.g.*, using recombinant antibody binding site expressing phage display libraries, in addition to the traditional *in vivo* methods using animals. See, *e.g.*, Huse (1989) *Science* 246:1275; Ward (1989) *Nature* 341:544;

Hoogenboom (1997) Trends Biotechnol. 15:62-70; Katz (1997) Annu. Rev. Biophys. Biomol. Struct. 26:27-45. Human antibodies can be generated in mice engineered to produce only human antibodies, as described by, e.g., U.S. Patent No. 5,877,397; 5,874,299; 5,789,650; and 5,939,598. B-cells from these mice can be immortalized using standard techniques (e.g., by fusing with an immortalizing cell line such as a myeloma or by manipulating such B-cells by other techniques to perpetuate a cell line) to produce a monoclonal human antibody-producing cell. See, e.g., U.S. Patent No. 5,916,771; 5,985,615.

TABLE 3

>Rv0002 dnaN DNA polymerase III, b-subunit TB.seq 2052:3257 MW:42114

>emb|AL123456|MTBH37RV:2052-3260, dnaN SEQ ID NO:1

ATGGACGCGGCTACGACAAGAGTTGGCCTCACCGACTTGACGTTTCGTTTGCTACGAGAGTCTT
TCGCCGATGCGGTGTCGTGGGTGGCTAAAAATCTGCCAGCCAGGCCCGCGGTGCCGGTGCTCT
CCGGCGTGTTGTTGACCGGCTCGGACAACGGTCTGACGATTTCCGGATTCGACTACGAGGTTTC
CGCCGAGGCCAGGTTGGCGCTGAAATTGTTTCTCCTGGAAGCGTTTTAGTTTCTGGCCGATTG
TTGTCCGATATTACCGGGCGTTGCCTAACAAGCCCGTAGACGTTTCATGTGAAGGTAACCGGG
TCGCATTGACCTGCGGTAACGCCAGGTTTTCGCTACCGACGATGCCAGTCGAGGATTATCCGAC
GCTGCCGACGCTGCCGGAAGAGACCGGATTGTTGCCTGCGGAATTATTCGCCGAGGCAATCAG
TCAGGTCGCTATCGCCGCCGGCCGGGACGACACGTTGCCTATGTTGACCGGCATCCGGGTCCA
AATCCTCGGTGAGACGGTGTTTTGGCCGCTACCGACAGTTTTGCCTGGCTGTTTCGAGAACTG
AAGTGGTCGGCGTCGTCGCCAGATATCGAAGCGGCTGTGCTGGTCCCGGCCAAGACGCTGGC
CGAGGCCGCCAAGCGGGCATCGGCCGGCTCTGACGTTTCGTTTGCTGTTGGGTACTGGGCCGG
GGGTGGGCAAGGATGGCCTGCTCGGTATCAGTGGGAACGGCAAGCGCAGCACCACGCGACTT
CTTGATGCCGAGTTCCCGAAGTTTCGGCAGTTGCTACCAACCGAACACACCGCGGTGGCCACC
ATGGACGTGGCCGAGTTGATCGAAGCGATCAAGCTGGTTGCGTTGGTAGCTGATCGGGGCGCG
CAGGTGCGCATGGAGTTCGCTGATGGCAGCGTGCGGCTTTCTGCGGGTGCCGATGATGTTGGA
CGAGCCGAGGAAGATCTTGTTGTTGACTATGCCGGTGAACCATTGACGATTGCGTTTAACCCAA
CCTATCTAACGGACGGTTTGAGTTGCTTGCCTCGGAGCGAGTGTCTTCGGGTTTACGACTGC
GGGTAAGCCTGCCTTGCTACGTCCGGTGTCCGGGGACGATCGCCCTGTGGCGGGTCTGAATGG
CAACGGTCCGTTCCCGGCGGTGTCGACGGACTATGTCTATCTGTTGATGCCGGTTCGGTTGCCG
GGCTGA

>Rv0003 recF DNA replication and SOS induction TB.seq 3280:4434 MW:42181

>emb|AL123456|MTBH37RV:3280-4437, recF SEQ ID NO:2

GTGTACGTCCGTCATTTGGGGCTGCGTGACTTCCGGTCTGGGCATGTGTAGATCTGGAATTGC
ATCCAGGGCGGACGGTTTTTGTGGGCCTAACGGTTATGGTAAGACGAATCTTATTGAGGCACT

GTGGTATTCGACGACGTTAGGTTGCGACCGCGTTAGCGCCGATTTGCCGTTGATCCGGGTAGGT
ACCGATCGTGCGGTGATCTCCACGATCGTGGTGAACGACGGTAGAGAATGTGCCGTGACCTC
GAGATCGCCACGGGGCGAGTCAACAAAGCGCGATTGAATCGATCATCGGTCCGAAGTACACGT
GATGTGGTCCGAGTGCTTCGAGCTGTGTTGTTTCCCCCTGAGGATCTGGGGTTGGTTCGTGGG
5 GATCCCGCTGACCGGGCGGCGCTATCTGGATGATCTGGCGATCGTGGTAGGCCCTGCGATCGCT
GCGGTACGAGCCGAATATGAGAGGGTGTTGCGCCAGCGGACGGCGTTATTGAAGTCCGTACCT
GGAGCACGGTATCGGGGTGACCGGGGTGTGTTTGACACTCTTGAGGTATGGGACAGTCGTTTG
GCGGAGCACGGGGCTGAACTGGTGGCCGCCCGCATCGATTTGGTCAACCAGTTGGCACCGGA
AGTGAAGAAGGCATACCAGCTGTTGGCGCCGGAATCGCGATCGGCGTCTATCGGTTATCGGGC
10 CAGCATGGATGTAACCGGTCCCAGCGAGCAGTCAGATATCGATCGGCAATTGTTAGCAGCTCGG
CTGTTGGCGGCGCTGCGGGCCCCGTCGGGATGCCGAACTCGAGCGTGGGGTTTGTCTAGTTGGT
CCGCACCGTGACGACCTAATACTGCGACTAGGCGATCAACCCGCGAAAGGATTTGCTAGCCATG
GGGAGGCGTGGTCGTTGGCGGTGGCACTGCGGTTGGCGGCCTATCAACTGTTACGCGTTGATG
GTGGTGAGCCGGTGTGTTGCTCGACGACGTGTTGCCGAACCTGGATGTCATGCGCCGTGAG
15 CGTTGGCGACGGCGGCGGAGTCCGCCGAACAGGTGTTGGTGACTGCCGCGGTGCTCGAGGAT
ATCCCGCCGGCTGGGACGCCAGGCGGGTGCACATCGATGTGCGTGCCGATGACACCGGATC
GATGTCGGTGGTCTGCCATGA

>Rv0005 gyrB DNA gyrase subunit B TB.seq.5123:7264 MW:78441
20 >emb|AL123456|MTBH37RV:5123-7267, gyrB SEQ ID NO:3
ATGGGTAAAAACGAGGCCAGAAGATCGGCCCTGGCGCCCGATCACGGTACAGTGGTGTGCGAC
CCCCTGCGGCGACTCAACCGCATGCACGCAACCCCTGAGGAGAGTATTCGGATCGTGGCTGCC
CAGAAAAAGAAGGCCCAAGACGAATACGGCGCTGCGTCTATCACCATTCTCGAAGGGCTGGAG
GCCGTCCGCAAACGTCCCGGCATGTACATTGGCTCGACCGGTGAGCGCGGTTTACACCATCTC
25 ATTTGGGAGGTGGTCGACAACGCGGTCGACGAGGCGATGGCCGGTTATGCAACCACAGTGAAC
GTAGTGCTGCTTGAAGATGGCGGTGTGAGGTGCGCGACGACGGCCGCGGCATTCCGGTCGC
CACCCACGCCTCCGGCATAACCGACCGTCGACGTGGTGATGACACAACCTACATGCCGGCGGCAA
GTTGACTCGGACGCGTATGCGATATCTGGTGGTCTGCACGGCGTCGGCGTGTGCGTGGTTAA
CGCGCTATCCACCCGGCTCGAAGTCGAGATCAAGCGCGACGGGTACGAGTGGTCTCAGGTTTA
30 TGAGAAGTCGGAACCCCTGGGCCTCAAGCAAGGGGCGCCGACCAAGAAGACGGGGTCAACGG
TGCGGTTCTGGGCCGACCCCGCTGTTTTCGAAACCACGGAATACGACTTCGAAACCGTCGCCC
GCCGGCTGCAAGAGATGGCGTTCCTCAACAAGGGGCTGACCATCAACCTGACCGACGAGAGGG
TGACCCAAGACGAGGTGTCGACGAAGTGGTCAGCGACGTGCGCGAGGCGCCGAAGTCGGCA
AGTGAACGCGCAGCCGAATCCACTGCACCGCACAAAGTTAAGAGCCGCACCTTTCACTATCCGG
35 GTGGCCTGGTGGACTTCGTGAAACACATCAACCGCACCAAGAACGCGATTATAGCAGCATCGT
GGACTTTTCCGGCAAGGGCACCGGGCACGAGGTGGAGATCGCGATGCAATGGAACGCCGGGT
ATTCGGAGTCGGTGCACACCTTCGCCAACACCATCAACACCCACGAGGGCGGCACCCACGAAG

AGGGCTTCCGCAGCGCGCTGACGTCGGTGGTGAACAAGTACGCCAAGGACCGCAAGCTACTGA
AGGACAAGGACCCCCAACCTCACCAGGTGACGATATCCGGGAAGGCCTGGCCGCTGTGATCTCGG
TGAAGGTCAGCGAACCGCAGTTCGAGGGCCAGACCAAGACCAAGTTGGGCAACACCGAGGTCA
AATCGTTTGTGCAGAAGGTCTGTAACGAACAGCTGACCCACTGGTTTGAAGCCAACCCACCGA
5 CGCGAAAGTCGTTGTGAACAAGGCTGTGTCTCGGCGCAAGCCCGTATCGCGGCACGTAAGGC
ACGAGAGTTGGTGCGGCGTAAGAGCGCCACCGACATCGGTGGATTGCCCCGGAAGCTGGCCG
ATTGCCGTTCCACGGATCCGCGCAAGTCCGAACGTGTATGTCGTAGAAGGTGACTCGGCCGGCG
GTTCTGCAAAAAGCGGTGCGGATTCGATGTTCCAGGCGATACTTCCGCTGCGCGGCAAGATCAT
CAATGTGGAGAAAGCGCGCATCGACCGGGTGCTAAAGAACACCGAAGTTCAGGCGATCATCAC
10 GCGCGCTGGGCACCGGGATCCACGACGAGTTGATATCGGCAAGCTGCGCTACCACAAGATCGT
GCTGATGGCCGACGCCGATGTTGACGGCCAACATATTTCCACGCTGTTGTTGACGTTGTTGTC
CGGTTTCATGCGGCCGCTCATCGAGAACGGGCATGTGTTTTTGGCACAACCGCCGCTGTACAAAC
TCAAGTGGCAGCGCAGTGACCCGGAATTCGCATACTCCGACCGCGAGCGCGACGGTCTGCTGG
AGGCGGGGCTGAAGGCCGGGAAGAAGATCAACAAGGAAGACGGCATTACGCGGTACAAGGGT
15 CTAGGTGAAATGGACGCTAAGGAGTTGTGGGAGACCACCATGGATCCCTCGGTTCTGTGTTGC
GTCAAGTGACGCTGGACGACGCCGCCGCCGCCGACGAGTTGTTCTCCATCCTGATGGGCGAGG
ACGTCGACGCGCGCGCGCAGCTTTATCACCCGCAACGCCAAGGATGTTGCGTTCCTGGATGTCTA
A

20 >Rv0006 gyrA DNA gyrase subunit A TB.seq 7302:9815 MW:92276
>emb|AL123456|MTBH37RV:7302-9818, gyrA SEQ ID NO:4
ATGACAGACACGACGTTGCCGCCTGACGACTCGCTCGACCGGATCGAACCAGGTTGACATCGAG
CAGGAGATGCAGCGCAGCTACATCGACTATGCGATGAGCGTGATCGTCGGCCGCGCGCTGCCG
GAGGTGCGCGACGGGCTCAAGCCCGTGCATCGCCGGGTGCTCTATGCAATGTTGATTCCGGC
25 TTCCGCCCGGACCGCAGCCACGCCAAGTCGGCCCGGTGCGTTGCCGAGACCATGGGCAACTA
CCACCCGCACGGCGACGCGTCGATCTACGACAGCCTGGTGCGCATGGCCCAGCCCTGGTCGC
TGCGCTACCCGCTGGTGACGGCCAGGGCAACTTCGGCTCGCCAGGCAATGACCCACCGGCG
GCGATGAGGTACACCGAAGCCCGGCTGACCCCGTTGGCGATGGAGATGCTGAGGGAAATCGAC
GAGGAGACAGTCGATTTTCATCCCTAACTACGACGGCCGGGTGCAAGAGCCGACGGTGCTACCC
30 AGCCGGTTCCCCAACCTGCTGGCCAACGGGTGAGGCGGCATCGCGGTGCGCATGGCAACCAAT
ATCCCGCCGCACAACCTGCGTGAGCTGGCCGACGCGGTGTTCTGGGCGCTGGAGAATCACGAC
GCCGACGAAGAGGAGACCCTGGCCGCGGTGATGGGGCGGGTTAAAGGCCCGGACTTCCCGAC
CGCCGGACTGATCGTCGGATCCCAGGGCACCGCTGATGCCTACAAAACCTGGCCGCGGCTCCAT
TCGAATGCGCGGAGTTGTTGAGGTAGAAGAGGATCCCGCGGTGCTACCTCGCTGGTGATCAC
35 CGAGTTGCCGTATCAGGTCAACCACGACAACTTCATCACTTCGATCGCCGAACAGGTCCGAGAC
GGCAAGCTGGCCGGCATTTCACACATTGAGGACCAGTCTAGCGATCGGGTCGGTTTACGCATC
GTCATCGAGATCAAGCGCGATGCGGTGGCCAAGGTGGTGATCAATAACCTTTACAAGCACACCC

AGCTGCAGACCAGCTTTGGCGCCAACATGCTAGCGATCGTCGACGGGGTGCCGCGCACGCTGC
GGCTGGACCAGCTGATCCGCTATTACGTTGACCACCAACTCGACGTCATTGTGCGGCGCACAC
CTACCGGCTGCGCAAGGCAAACGAGCGAGCCCACATTCTGCGCGGCCTGGTTAAAGCGCTCGA
CGCGCTGGACGAGGTCATTGCACTGATCCGGGCGTCGGAGACCGTCGATATCGCCCGGGCCG
5 GACTGATCGAGCTGCTCGACATCGACGAGATCCAGGCCAGGCAATCCTGGACATGCAGTTGC
GGCGCCTGGCCGCACTGGAACGCCAGCGCATCATCGACGACCTGGCCAAAATCGAGGCCGAG
ATCGCCGATCTGGAAGACATCCTGGCAAACCCGAGCGGCAGCGTGGGATCGTGCGCGACGAA
CTCGCCGAAATCGTGACAGGCACGGCGACGACCGGCGTACCCGGATCATCGCGGCCGACGG
AGACGTCAGCGACGAGGATTTGATCGCCCGCGAGGACGTCGTTGTCACTATCACCGAAACGGG
10 ATACGCCAAGCGCACCAAGACCGATCTGTATCGCAGCCAGAAACGCGGCGGCAAGGGCGTGCA
GGGTGCGGGGTTGAAGCAGGACGACATCGTCGCGCACTTCTTCGTGTGCTCCACCCACGATTT
GATCCTGTTCTTACCACCCAGGGACGGGTTTATCGGGCCAAGGCCTACGACTTGCCCGAGGC
CTCCCGGACGGCGCGCGGGCAGCACGTGGCCAACCTGTTAGCCTTCCAGCCCGAGGAACGCA
TCGCCCAGGTCATCCAGATTCGCGGCTACACCGACGCCCCGTACCTGGTGCTGGCCACTCGCA
15 ACGGGCTGGTGAAAAAGTCCAAGCTGACCGACTTCGACTCCAATCGCTCGGGCGGAATCGTGG
CGGTCAACCTGCGCGACAACGACGAGCTGGTTCGGTGCGGTGCTGTGTTGCGCCGGCGACGAC
CTGCTGCTGGTCTCGGCCAACGGGCAGTCCATCAGGTTCTCGGCGACCGACGAGGCGCTGCG
GCCAATGGGTCTGCCACCTCGGGTGTGCAGGGCATGCGGTTCAATATCGACGACCGGCTGCT
GTCGCTGAACGTCGTGCGTGAAGGCACCTATCTGCTGGTGGCGACGTCAGGGGGCTATGCGAA
20 ACGTACCGCGATCGAGGAATACCCGGTACAGGGCCGCGGCGGTAAAGGTGTGCTGACGTCAT
GTACGACCGCCGGCGCGGCAGGTTGGTTGGGGCGTTGATTGTCGACGACGACAGCGAGCTGT
ATGCCGTCACTTCCGGCGGTGGCGTGATCCGCACCGCGGCACGCCAGGTTGCAAGGCGGGA
CGGCAGACCAAGGGTGTTCCGGTTGATGAATCTGGGCGAGGGCGACACACTGTTGGCCATCGCG
CGCAACGCCGAAGAAAGTGCGACGATAATGCCGTGGACGCCAACGGCGCAGACCAGACGGG
25 CAATTAA

>Rv0014c pknB serine-threonine protein kinase TB.seq 15593:17470 MW:66511

>emb|AL123456|MTBH37RV:c17470-15590, pknB SEQ ID NO:5

ATGACCACCCCTTCCACCTGTCCGACCGCTACGAACTTGCGCAAATCCTTGGAATTTGGGGGCA
30 TGTCCGAGGTCCACCTGGCCCCGCGACCTCCGGTTGCACCGCGACGTTGCGGTCAAGGTGCTGC
GCGCTGATCTAGCCCGCGATCCAGTTTTTACCTTCGCTTCCGGCGTGAGGCGCAAAACGCCG
CGGCATTGAACCACCCTGCAATCGTCGCGGTCTACGACACCGGTGAAGCCGAAACGCCCGCCG
GGCCATTGCCCTACATCGTCATGGAATACGTCGACGCGGTTACCCTGCGCGACATTGTCCACAC
CGAAGGGCCGATGACGCCCAAACGCGCCATCGAGGTCATCGCCGACGCTGCCAAGCGCTGA
35 ACTTCAGTCATCAGAACGGAATCATCCACCGTGACGTCAAGCCGGCGAACATCATGATCAGCGC
GACCAATGCAGTAAAGGTGATGGATTTCCGCATCGCCCGCGCCATTGCCGACAGCGGCAACAG
CGTGACCCAGACCGCAGCAGTGATCGGCACGGCGCAGTACCTGTCACCCGAACAGGCCCGGG

GTGATTCCGTCGACGCCCGATCCGATGTCTATTCTTGGGCTGTGTTCTTTATGAAGTCCTCACC
GGGGAGCCACCTTTACCGGGCGACTCACCCGTCTCGGTTGCCTACCAACATGTGCGCGAAGAC
CCGATCCCACCTTCGGCGCGGCACGAAGGCCTCTCCGCCGACCTGGACGCCGTCGTTCTCAAG
GCGCTGGCCAAAAATCCGGAAAAACCGCTATCAGACAGCGGCGGAGATGCGCGCCGACCTGGTC
5 CCGCTGCACAACGGTGAGCCGCCCGAGGCGCCCAAAGTGCTCACCAGATGCCGAGCGGACCTC
GCTGCTGTCGTCTGCGGCCGGCAACCTTAGCGGTCCGCGCACCGATCCGCTACCACGCCAGGA
CTTAGACGACACCGACCGTGACCGCAGCATCGGTTCCGGTGGGCCGTTGGGTTGCGGTGGTCCG
CGTGCTCGCTGTGCTGACCGTCGTGGTAACCATCGCCATCAACACGTTCCGGCGGCATCACCCG
CGACGTTCAAGTTCCCGACGTTCCGGGTCAATCCTCCGCCGACGCCATCGCCACACTGCAAAA
10 CCGGGGCTTCAAATCCGCACCTTGACAGAGCCGGACTCGACAATCCACCGGACCACGTTAT
CGGCACCGACCCGGCCGCCAACACGTCCGTGAGTGACGGCGACGAGATCACAGTCAACGTGT
CCACCGGACCCGAGCAACGCGAAATACCCGACGTCTCCACGCTGACATACGCCGAAGCGGTCA
AGAACTGACTGCCGCCGGATTCCGGCCGCTTCAAGCAAGCGAATTCGCCGTCCACCCCGGAAC
TGGTGGGCAAGGTCATCGGGACCAACCCGCCAGCCAACCAGACGTCCGGCCATACCAATGTGG
15 TCATCATCATCGTTGGCTCTGGTCCGGCGACCAAAGACATTCCCGATGTCGCGGGCCAGACCGT
CGACGTGGCGCAGAAGAACCTCAACGTCTACGGCTTACCAAATTCAGTCAGGCCTCGGTGGA
CAGCCCCCGTCCCGCCGGCGAGGTGACCGGCACCAATCCACCCGCAGGCACCACAGTTCCGG
TCGATTAGTCATCGAACTACAGGTGTCCAAGGGCAACCAATTCGTCATGCCCGACCTATCCGG
CATGTTCTGGGTCGACGCCGAACCACGATTGCGCGCGCTGGGCTGGACCGGGATGCTCGACAA
20 AGGGGCCGACGTGACGCCGGTGGCTCCCAACACAACCGGGTCGTCTATCAAAACCCGCCGG
CGGGGACCGGCGTCAACCGGGACGGCATCATCACGCTGAGGTTCCGGCCAGTAG

>Rv0016c pbpA TB.seq 18762:20234 MW:51577

>emb|AL123456|MTBH37RV:c20234-18759, pbpA SEQ ID NO:6

25 ATGAACGCCTCTCTGCGCCGAATATCGGTGACCGTGATGGCGTTGATCGTGTTGCTACTGCTCA
ACGCGACCATGACGCAGGTCTTACCGCCGACGGGCTGCGTGCCGATCCCCGCAACCAGCGA
GTGTTGCTCGACGAGTATTCACGGCAGCGCGGCCAGATCACCGCTGGTGGCCAACTGCTGGCG
TACTCGGTAGCCACCGACGGCCGCTTTCGTTTCCTGCGGGTCTATCCCAATCCTGAGGTGTACG
CGCCGGTTACCGGCTTCTACTCCCTGCGCTATTCCAGCACCGCCCTAGAACGAGCCGAGGACC
30 CGATATTGAACGGGTCCGACCGCCGTCTGTTCCGGCCGCCGGTGGCCGACTTCTTACCGGTC
GCGACCCACGCGGCGGTAATGTCGATACCACGATCAACCCGCGCATTACGCAAGCCGGCTGGG
ACGCGATGCAGCAAGGCTGCTACGGGGCCCTGTAAGGGAGCGGTGGTCGCCCTTGAGCCATCAA
CCGGCAAGATTTTGGCGTTGGTGTCTTCTCCGTCTACGACCCCAACCTGCTGGCGTCGCATAA
CCCCGAGGTGCAGGCGCAAGCCTGGCAGCGGCTTGGCGACAATCCCGCCTCTCCACTGACCAA
35 CCGTGCCATCTCTGAGACGTATCCACCGGGTTCGACTTTCAAAGTGATCACCCTGCGGCCGCG
CTGGCCGCCGGGGCCACCGAGACCGAACAGCTGACTGCGGCGCCACAATTCCGTTGCCAGG
CAGCACCGCCAGCTAGAGAACTACGGCGGTGCGCCGTGCGGGGACGAACCCACCGTGTGCG

TGCGTGAGGCATTCGTCAAATCATGCAACACCGCATTCGTCCAGCTGGGCATCCGCACCGGCG
CCGACGCCCTGCGCAGCATGGCGCGCGCGTTCGGTCTCGATAGCCACCGCGCCCAACTCCG
CTGCAAGTGGCGGAATCAACCGTCGGGCCTATCCCGGACAGCGCCGCACTAGGGATGACCAGT
ATCGGCCAAAAGGACGTTGCGCTGACCCCGCTAGCGAACGCAGAAATAGCCGCGACCATCGCA
5 AACGGCGGCATTACGATGAGGCCCTTATCTAGTCGGCAGCCTCAAGGGACCGGACCTAGCCAAT
ATCTCAACCACCGTCGGATACCAGCAGCGCCGCGCGGTGTCACCGCAGGTGCGCGCTAAGCTA
ACAGAGCTGATGGTCGGCGCCGAGAAAGTCGCACAGCAGAAAGGGGCAATCCCCGGCGTGCA
GATCGCATCCAAGACGGGCACCGCCGAACATGGCACCGACCCCTCGTCACACTCCACCGCACGC
TTGGTACATCGCCTTTGCGCCCGCACAAAGCGCCCAAGGTGGCTGTTGCCGTGCTGGTGGAGAA
10 CGGGGCTGATCGGCTGTCCGCCACCGGAGGTGCCCTCGCGGCACCGATCGGGCGGGCGGTG
ATCGAAGCCGCACTGCAGGGGAACCATGA

>Rv0017c rodA TB.seq 20234:21640 MW:50812

>emb|AL123456|MTBH37RV:c21640-20231, rodA SEQ ID NO:7

15 ATGACGACACGACTGCAAGCGCCGGTGGCCGTAACGCCCCCGTTGCCGACTCGGCGCAACGC
TGAACTGCTGCTGCTGTGCTTTGCCGCCGTAATCACGTTTGCCGCACTGCTGGTCGTGCAGGCC
AATCAAGACCAGGGGGTGCCCTGGGACTTGACTAGCTACGGACTGGCCTTCCTGACCCTGTTT
GGATCCGCGCATCTGGCCATCCGGCGCTTCGCCCCCTACACTGACCCGCTGTTGCTCCCGGTG
GTGGCACTGCTCAACGGACTTGGCCTGGTAATGATCCACCGCCTCGATCTGGTGGACAACGAG
20 ATCGGCGAGCATCGGCACCCAGCGCAAACCAGCAGATGCTGTGGACGCTGGTGGGCGTAGC
TGCTTCGCGCTCGTGGTGACCTTCCTCAAGGACCACCGACAGCTCGCACGCTACGGCTACATT
TGCGGGCTCGCGGGTCTGGTTTTCTTGGCAGTTCCCGCGCTGCTCCCGGCAGCACTGTCCGAA
CAGAACGGCGCCAAGATCTGGATCCGGTTGCCCGGCTTCTCGATTCAACCCGCCGAATTTTCAA
AGATTCTGCTGCTGATCTTCTTTTGGCGGTACTGGTGGCCAAACGCGGCCTGTTACACGCGC
25 CGGCAAACATTTGCTCGGAATGACCCTGCCGCGCCCGCGAGACCTCGCGCCACTGTTGGCAGC
CTGGGTATCTCGGTGGGTGTGATGGTCTTCGAGAAAGACCTCGGCGCTTCGCTGCTGCTGTAC
ACATCGTTTCTGGTGGTGGTTTACCTCGCCACCCAGCGGTTTCAGTTGGGTCGTCATCGGCCTGA
CTCTGTTGCGCGCAGGAACCTTGGTGGCGTACTTCATTTTTGAGCACGTCCGGCTCCGCGTACA
GACCTGGCTGGATCCGTTTCGCAGATCCAGACGGCACCGGATATCAGATCGTGCACTCGCTTTTC
30 AGCTTCGCTACAGGCGGTATCTTCGGCACCGGGCTCGGTAATGGTCAACCCGACACCGTGCCC
GCGGCATCCACCGATTTTCATCATCGCCGCGTTCGGCGAAGAGCTTGGGTTGGTGGGCTTGACG
GCCATCCTGATGCTCTACACCATCGTGATCATCCGGGGTTTGCGCACGGCCATCGCCACCCGC
GATAGCTTCGGCAAGCTGCTGGCCGCGCGCCTCTCATCGACGCTAGCCATTAGCTGTTTCATCG
TCGTGGCGGGTGTGACCCGACTCATTCCGCTGACCGGGTTGACCACACCGTGGATGTCTACG
35 GCGGGTCTTCACTGCTGGCCAACTACATATTGCTGGCCATCCTGGCACGCATCTCGCACGGAGC
CCGCCGCCCACTGCGCACCCGCGCCACGAAATAAGTCGCCGATTACGGCGGCCGGCACCGAGG
TCATCGAACGCGTATGA

>Rv0018c ppp TB.seq 21640:23181 MW:53781

>emb|AL123456|MTBH37RV:c23181-21637, ppp SEQ ID NO:8

GTGGCGCGCGTGACCCCTGGTCCTGCGATACGCGGCGCGCAGCGATCGCGGCTTGGTACGCGC
5 CAACAACGAAGACTCGGTCTACGCTGGGGCACGGCTATTGGCCCTGGCCGACGGCATGGGTG
GGCATGCGGCCGGCGAGGTGGCGTCCCAGTTGGTGATTGCCGCATTGGCCCATCTCGATGACG
ACGAGCCCGGTGGCGATCTGCTGGCCAAGCTGGATGCCGCGGTGCGCGCCGGCAACTCGGCT
ATCGCAGCGCAAGTCGAGATGGAGCCCGATCTCGAAGGCATGGGTACCACGCTCACC GCAATC
CTGTTGCGGGCAACCGGCTCGGCCTGGTGCATATCGGTGACTCGCGCGGTTACCTGCTGCGC
10 GACGGTGAGCTGACGCAGATCACCAAGGACGACACGTTTGTCCAAACGCTGGTCGACGAAGGC
CGGATCACCCCGGAGGAGGCGCACAGCCACCCGCAACGCTCGTTGATCATGCGGGCGTTGAC
CGGCCATGAGGTGGAACCGACGCTGACCATGCGAGAAGCCGCGCCGGTGATCGTTACCTGCT
GTGCTCGGACGGGTTGTCCGATCCGGTTAGCGATGAACTATCCTCGAGGCCCTGCAGATCCC
CGAGGTTGCCGAGAGCGCTCACCGCCTCATTGAACTGGCGCTGCGCGGCGGGCGGCCCCGACA
15 ACGTCACTGTGTCGTCGCCGACGTCGTCGACTACGACTACGGCCAGACCCAACCGATTCTGG
CCGGGGCGGTCTCAGGCGACGACGACCAACTGACCCTGCCCAACACCGCCGCGGCGGGGCC
TCTGCCATCAGCCAGCGCAAGGAGATCGTTAAACGCGTTCCGCCACAGGCCGATACATTAGTC
GGCCACGGTGGTCGGGCGGACGGCTAGCATTGTTGTCGCACTGGTGACCGTGCTGATGACTG
CGGGCCTGCTCATTGGTCGCGCGATCATCCGCAGCAACTACTACGTAGCGGACTACGCCGGCA
20 GCGTGTCCATCATGCGGGGGATTCAAGGGTCGCTACTGGGCATGTCCCTGCACCAGCCTTACC
TGATGGGCTGCCCTAGCCCGCGTAACGAGCTGTGCGAGATCAGCTACGGACAGTCTGGGGGCC
CTCTCGACTGCCATCTGATGAACTGGAGGATCTGCGACCGCCGAGCGCGCACAGGTTCCGG
CCGGTCTCCCGGCCGGCACTCTCGATGACGCCATCGGGCAGTTGCGCGAACTGGCGGGCCAACT
CCCTGCTGCCGCCTTGCCCGGCGCGCGTGCACGTCCCCGCCCGGGCGCCCCGGCCCCACCC
25 ACCACCAGCGAGACAACCGAACC AAACGTACCTCCTCGCCAGCCTCTCCATCACCCACCACCT
CCGCGCCGGCCCCCACC GAACTACTCCTGCCATCCCCACGAGTGCCCTCCCCGGCAGCGCCC
GCGTCGCCGCGGACGCCTTGCCCCGTACCAGCTCGCCGACGATGGCCGCACTTCCGCCACC
CCCGCCTCAGCCGGGCATCGACTGCCGGGCGGCGGCATGA

30 >Rv0019c - TB.seq 23273:23737 MW:17153

>emb|AL123456|MTBH37RV:c23737-23270, Rv0019c SEQ ID NO:9

ATGCAGGGGTTGGTACTGCAACTGACGCGTGCCGGATTCTTGATGTTGTTGTGGGTATTCATCT
GGTCCGTGCTACGGATCTTGAAGACCGACATTTATGCGCCGACCGGCGCGGTATGATGCGCC
GCGGCCTGGCGCTGCGAGGGACGCTCTTAGGCGCGCGTCAGCGCCGGCACGCTGCACGCTAC
35 CTGGTGGTGACCGAAGGTGCGTTGACTGGCGCGCGTATCACGCTGAGCGAACAGCCGGTGTG
ATCGGGCGCGCCGACGACTCGACCCTGGTGCTGACCGACGACTACGCCTCGACGCGGCACGC
TCGGCTGTCTATGCGCGGCTCCGAGTGGTACGTGGAAGATCTAGGATCGACCAACGGCACTTA

CCTGGACAGGGCGAAGGTGACGACTGCGGTACGAGTTCCGATCGGAACGCCGGTTCGCATCG
GCAAAACTGCAATCGAGTTGCGCCCGTGA

>Rv0020c - TB.seq 23864:25444 MW:56881

5 >emb|AL123456|MTBH37RV:c25444-23861, Rv0020c SEQ ID NO:10
ATGGGTAGCCAGAAAAGGCTGGTTCAGCGCGTTGAGCGCAAACCTCGAGCAGACGGTTGGCGAT
GCGTTTGCCCGCATCTTTGGAGGCTCGATCGTCCCGCAAGAGGTGGAAGCCCTGCTGCGCCGC
GAGGCGGCCGACGGCATCCAGTCGCTGCAGGGAAATCGCCTTTTGCGCCCAACGAATACATC
ATTACCCTCGGTGTGCACGACTTTGAGAAGTTGGGCGCTGATCCTGAGCTGAAGTCAACCGGTT
10 TTGCTCGGGACTTGGCGGACTATATCCAAGAACAGGGGTGGCAAACGTATGGTGATGTGGTCGT
CCGATTCGAGCAGTCGTGGAACCTGCATACCGGCCAGTTCCGCGCCCGCGGCACTGTAAACCC
CGACGTTGAGACCCACCCGCCGGTCATCGATTGCGCCCGGCCACAATCAAACCACGCGTTTGG
CGCAGAACCAGGAGTAGCACCAATGAGTGACAATTCGAGCTACCGTGCGGGTCAGGGGCAGGG
GCGTCCCGACGAGTATTACGACGACCGCTATGCGCGTCCGCAAGAGGATCCGCGTGGTGGCCC
15 GGATCCGCAAGGCGGATCTGACCCCGCGGGGGGTATCCACCCGAGACGGGCGGCTACCCGCG
CCCAGCCGGGCTACCCACGCCCGCGCCACCCGGACAGGGCGACTACCCCGAGCAAATCGGG
TACCCCGACAGGGCGGTTACCCCGAGCAACGCGGTTACCCCGAGCAACGCGGCTACCCCGA
CCAGCGCGGGTACCAGGACCAGGGTCGAGGCTACCCCGACCAAGGGCAGGGGGGCTATCCGCG
CGCCCTACGAGCAACGCCCTCCTGTTTCTCCCGGCCCGGCTGCCGGCTACGGCGCTCCCGGCT
20 ACGACCAGGGCTATCGCCAAAGCGGCGGCTACGGCCCTTCACCCGGTGCGGGCCAGCCCGGC
TACGGCGGGTACGGGGAGTACGGGCGTGGCCCGGCTCGCCACGAGGAGGGCAGCTATGTGCC
CTCTGGCCCTCCGGGCCCGCCCGAGCAACGACCGGCTTACCCCGACCAAGGCGGTTACGACC
AGGGCTACCAGCAAGGCGCCACGACATACGGCCGGCAAGACTATGGCGGCGGCGCTGACTAC
ACCCGCTACACCGAATCCCGCGGGTCCCGGGATACGCTCCTCAGGGTGGCGGGTACGCCGA
25 ACCCGCCGGCCGAGACTACGACTACGGCCAATCAGGCGCTCCGGACTACGGTCAGCCAGCGC
CCGGTGGCTACAGCGGTTACGGGCAGGGCGGCTATGGGTCCGCCGGAACGTGGTTACGCTG
CAGCTCGACGACGGCAGCGGACGCACTTACCAGCTCCGCGAGGGCTCCAACATCATCGGTGCG
GGACAGGACGCCAGTTCCGGCTGCCCGACACCGGTGTGTACGCCGTCACCTTGAGATCCG
GTGGGACGGGCAGGTGCGATTGCTCGCAGACCTGAACTCCACCAACGGCACCACTGTAAACAA
30 TGCACCGGTACAGGAGTGGCAGTTGGCCGACGGTGATGTGATCCGCTTGGGACACTCCGAGAT
CATCGTCCGCATGCACTGA

>Rv0032 bioF2 C-terminal similar to B. subtilis BioF TB.seq 34295:36607 MW:86245

>emb|AL123456|MTBH37RV:34295-36610, bioF2 SEQ ID NO:11

35 ATGCCCACTGGCTTGGGCTATGACTTTCTGCGCCCTGTCGAGGACTCGGGGATCAACGACCTGA
AGCACTATTACTTCATGGCGGATTTGGCCGATGGGCAACCGCTAGGCCGGGCAAACCTCTATAG
CGTCTGTTTCGACCTGGCCACCACCGACCGCAAGCTCACTCCGGCCTGGCGAACGACCATCAA

ACGGTGGTTTCCGGGGTTTATGACCTTCCGTTTCCTCGAGTGCGGGTTGCTCACCATGGTGAGC
AACCCGCTGGCGTTGCGGTCCGACACCGACTTGAGCGGGTATTGCCTGTGCTGGCCGGCCAG
ATGGACCAGTTGGCGCATGACGACGGGTCGGATTTCTTGATGATCCGGGACGTGGACCCGGAA
CACTACCAGCGATACCTTGACATCCTGCGCCCCGTTGGGCTTTTCGGCCTGCGCTGGGCTTTTCCC
5 GGGTAGACACGACCATCAGCTGGTCGAGCGTGGAAGAGGCACTGGGCTGCCTGTCTCACAAAA
GGCGCCTGCCGTTGAAGACGTGCTGGAGTTTCGTGAGCGGTTGGTATCGAGGTGAGGAAC
TCGACGAGTATGCCGAGCATGCGCCGGTATTGGCCCGGCTTTGGCGCAACGTCAAGACGGAGG
CAAAGGATTACCAGCGCGAGGACCTGAACCCTGAGTTCTTCGCGGCGTGTTCTCGGCATCTGCA
TGGACGTAGCAGACTGTGGTTGTTCCGCTACCAGGGCACGCCAATTGCCTTCTTTTTGAACGTTT
10 GGGGTGCGGATGAGAACTACATACTGCTTGAGTGGGGCATCGATCGTGATTTTGAACATTATAG
GAAGGCGAATCTGTACCGGGCGGCGCTGATGCTCAGCCTAAAAGATGCGATCAGCCGAGATAA
ACGGCGAATGGAAATGGGTATTACGAACTATTTACAAAACTTCGCATTCCGGGTGCCCCGAGTC
ATACCGACCATCTATTTCTGCGTCACAGCACGGATCCGGTGCATACGGCAACGTTAGCGCGAA
TGATGATGCACAATATTCAACGGCCAACGCTACCCGACGATATGTGCGAGGAATTCTGTCGCTG
15 GGAAGAGCGAATACGTCTGGACCAGGACGGGCTACCGAACACGATATCTTTGCAAGATCGAT
CGTCAGCACAAATACACGGGGCTCAAACCTCGGCGGAGTCTACGGTTTTTATCCCCGATTACCCG
GACCGCAGCGATCCACGGTCAAGGCCGCGGAGCTGGGCGAGATCGTGTTGCTGGGCACGAAC
TCGTATCTGGGCCTGGCCACCCATCCAGAGGTGGTGGAGGCCTCGGCGGAGGCCACGCGACG
GTACGGCACCGGCTGCTCGGGTTCGCCGTTGCTGAACGGCACGTTGGACTTGACGCTCTCGCT
20 TGAGCAGGAACTAGCCTGTTTTTGGGCAAACCCGCCGCCGTTGTGTGCTCCACCGGATATCAG
AGCAACCTGGCGGCGATCAGCGCGCTATGCGAATCCGGGGACATGATCATCCAAGACGCGCTG
AACCACCGCAGCCTGTTTCGACGCCGCCAGGTTGTCCGGGGCCGACTTCACCTTGACCGGCAC
AACGACATGGACCACCTGGCGCGGGTGCTACGCCGCACCGAGGGGCGCCGCCGGATCATCGT
CGTGGACGCGGTGTTTCAGCATGGAAGGCACCGTCGCCGACCTGGCCACCATCGCCGAGCTTG
25 CCGACCGGCACGGCTGCCGGGTCTATGTGGACGAGTCCCATGCGCTGGGCGTGCTCGGCCCC
GACGGGCGAGGAGCTTCGGCCGCGTTGGGTGCTTGGCGCGCATGGACGTGGTGATGGGCAC
GTTTCAGCAAATCCTTTGCCTCCGTGCGCGGGTTCATCGCCGGAGATCGGCCCGTCGTGGACTA
CATCCGGCACAACGGTTCAGGTCATGTGTTTTCCGCCAGCCTGCCGCCGGCCGCCGCGGCTGC
CACCCACGCGGCTCTGCGCGTCAGTCGGCGTGAACCCGACCGGCGGGCTCGGGTGCTGGCCG
30 CGGCCGAGTACATGGCCACCGGCCTGGCACGGCAGGGCTATCAGGCCGAGTATCACGGAACC
GCGATCGTGCCGGTGATCCTGGGCAACCCGACCGTGGCGCATGCGGGCTATCTGCGGCTGAT
GCGCTCCGGGGTGATGTGAACCCGGTGGCCCCCCCAGCCGTGCCGGAGGAGCGTTCCGGGAT
TCCGCACCAGCTACCTAGCCGACCAACGACAATCTGACCTCGACCGGGCCTTGACAGTGTTTGC
CGGCCTTGCCGAGGACCTGACCCCGCAAGGAGCCGCGCTATGA

35

>Rv0050 ponA1 TB.seq 53661:55694 MW:71119

>emb|AL123456|MTBH37RV:53661-55697, ponA SEQ ID NO:12

GTGGTGATCCTGTTGCCGATGGTCACCTTCACGATGGCCTACCTGATCGTCGACGTTCCC/AAGC
CAGGTGACATCCGTACCAACCAGGTCTCCACGATCCTTGCCAGCGACGGCTCGGAAATCGCCA
AAATTGTTCCGCCCGAAGGTAATCGGGTCGACGTCAACCTCAGCCAGGTGCCGATGCATGTGC
GCCAGGCGGTGATTGCGGCCGAAGACCGCAATTTCTATTCTGAATCCGGGATTCTCGTTACCCGG
5 CTTGCGCGGGCAGTCAAGAACAACCTGTTGCGCGGCGATCTGCAGGGCGGATCGACGATTAC
CCAGCAGTACGTCAAGAACGCGCTGGTCGGTTCGCGACAGCACGGGTGGAGCGGTCTGATGC
GCAAGGCCGAAAGAATTGGTCATCGCGACGAAGATGTCGGGGGAGTGGTCTAAAGACGATGTGC
TGCAGGCGTATCTGAACATCATCTACTTCGGCCGGGGCGCCTACGGCATTTCGGCGGCGTCCA
AGGCTTATTTGACAAAGCCCGTCGAGCAGCTGACCGTTGCCGAAGGGGCGTTGTTGGCAGCGC
10 TGATTGCGCGGCCTTCGACGCTGGACCCGGCGGTGACCCCGAAGGGGCCCATGCCCCGCTGG
AATTGGGTACTCGACGGCATGGTGAAACCAAGGCTCTCTCGCCGAATGACCGTGCGGCGCAG
GTGTTTCCCGAGACAGTGCCGCCCGATCTGGCCCGGCGAGAGAATCAGACCAAAGGACCCAAC
GGGCTGATCGAGCGGCAGGTGACAAGGGAGTTGCTCGAGCTGTTCAACATCGACGAGCAGACC
CTCAACACCCAGGGGCTGGTGGTCACCACCACGATTGATCCGCAGGCCCAACGGGCGGCGGA
15 GAAGGCGGTTGCGAAATACCTGGACGGGCGAGACCCCGACATGCGTGCCGCCGTGGTTTCCAT
CGACCCGCACAACGGGGCGGTGCGTGCGTACTACGGTGCGGACAATGCCAATGGCTTTGACTT
CGCTCAAGCGGGATTGCAGACTGGATCGTCGTTTAAGGTGTTTGCTCTGGTGGCCGCCCTTGAG
CAGGGGATCGGCCTGGGCTACCAGGTAGACAGCTCTCCGTTGACGGTCGACGGCATCAAGATC
ACCAACGTCGAGGGCGAGGGTTGCGGGACGTGCAACATCGCCGAGGCGCTCAAAATGTCGCT
20 GAACACCTCCTACTACCGGTGATGCTCAAGCTCAACGGCGGGCCACAGGCTGTGGCCGATGC
CGCGCACCAAGCGGCATTGCCTCCAGCTTCCCGGGCGTTGCGCACACGCTGTCCGAAGATGG
CAAGGGTGGACCGCCCAACAACGGGATCGTGTTGGGCCAGTACCAAACCCGGGTGATCGACAT
GGCATCGGCGTATGCCACGTTGGCCGCGTCCGGTATCTACCACCCGCCGATTTCGTACAGAA
25 GGTGGTCAGTGCCAACGGCCAGGTCTCTTCGACGCCAGCACCGCGGACAACACCGGCGATCA
GCGCATCCCCAAGGCGGTAGCCGACAACGTGACTGCGGCGATGGAGCCGATCGCAGGTTATTC
GCGTGGCCACAACCTAGCGGGTGGGCGGGATTTCGGCGGCCAAGACCGGCACTACGCAATTTG
GTGACACCACCGCGAACAAGACGCCTGGATGGTTCGGGTACACGCCGTCGTTGTCTACGGCTG
TGTGGGTGGGCACCGTCAAGGGTGACGAGCCACTGGTAACCGCTTCGGGTGCAGCGATTACG
GCTCGGGCCTGCCGTGCGACATCTGGAAGGCAACCATGGACGGCGCCTTGAAGGGCACGTCCG
30 AACGAGACTTTCCCCAAACCGACCGAGGTGCGTGGTTATGCCGGTGTGCCGCCGCCGCCGCCG
CCGCCGGAGGTACCACCTTCGAGACCGTCATCCAGCCACGGTCGAAATTGCGCCGGGGATT
ACCATCCCGATCGGTCCCCCGACCACCATTACCCTGGCGCCACCGCCCCCGGCCCGCCCGCT
GCGACTCCACGCGCCGCCGCGTGA

35 >Rv0051 - TB.seq 55694:57373 MW:61210

>emb|AL123456|MTBH37RV:55694-57376, Rv0051 SEQ ID NO:13

GTGACCGGCGCGCTGTCCCAAAGCAGCAACATCTCGCCACTTCCTTTGGCCGCCGATCTGCGG
AGCGCCGATAACCGCGATTGCCCCAGCCGCACCGACGTATTGGGTGCCGCTCTGGCGAATGTC
GTCGGTGGCCCGGTAGGCCGGCACGCGCTGATCGGCCGCACCCGGCTGATGACCCCGCTGCG
GGTGATGTTTGCAATCGCGTTGGTGTTCTGGCGCTCGGTTGGTCGACGAAAGCGGCCTGCTT
5 GCAGTCCACCGGAACCGGTCCAGGTGATCAGCGGGTGGCCAACTGGGATAACCAGCGTGCTTA
CTACCAGTTGTGCTACTCCGATACGGTGCCGCTCTATGGCGCTGAGTTATTGAGCCAAGGCAAG
TTTCCGTACAAATCAAGCTGGATCGAAACCGACAGCAACGGCACACCGCAGCTGCGCTACGAC
GGACAGATCGCGGTGCGCTATATGGAGTATCCGGTGCTGACTGGGATCTATCAGTACCTGTGCA
TGGCGATAGCCAAGACCTACACCGCGTTAAGCAAGGTGGCTCCCCTCCCGGTGGTTGCCGAAG
10 TGGTGATGTTCTTCAACGTCGCCGCGTTCGGTTTGGCGCTGGCGTGCGTGACAACCGTCTGGG
CGACCTCGGGCCTGGCCGGCCGCCGGATATGGGATGCGGCGCTGGTGGCCGCCCTACCGCTG
GTGATCTTTTCAATATTCACCAATTTTCGATGCGCTGGCAACGGGTTTGGCGACGAGTGGGCTGC
TGGCCTGGGCGCGGCGCAGACCGGTGCTTGCCGGTGTGCTGATCGGGTTGGGCTCCGCGGCG
AAACTGTATCCGCTGTTGTTCTTGTACCCGTTGTTGCTGCTGGGCATCCGGGCGCGGTGCGCTGA
15 ATGCTCTGGCCCGCACCATGGCGGCCGCGCGCGACCTGGTTGTTGGTGAATCTGCCGGTGA
TGCTGCTCTTTCCGCGCGGCTGGTCCGAGTTCTTCCGGCTCAACACCCGGCGCGGCGACGACA
TGGACTCGTTGTACAACGTCGTCAAGTCGTTACCGGCTGGCGTGCTTCGACCCACCCCTGG
GCTTCTGGGAGCCGCGCTGGTGTGAACACGGTTGTACGCTCTTGTTCGTGTTATGTTGTGC
GGCAATTGCTTACATCGCGCTCACCGCACCCACCGGCCGCGCGTGGCGCAGCTGACTTTCTT
20 GACGGTGGCCAGCTTCCTGTTGGTCAACAAGGTGTGGAGTCCCCAGTTCTCGCTTTGGCTGGTG
CCGCTGGCCGTGCTGGCTTTGCCGCACCGCCGGATCTTGCTGGCGTGATGACGATCGACGCG
TTGGTGTGGGTGCCGCGGATGTACTACCTATACGGCAACCCGAGCCGCTCGCTGCCCGAGCAG
TGGTTACACACGACGGTGTTGCTGCGTGACATCGCCGTGATGGTGCTGTGCGGACTGGTGGTC
TGGCAGATCTACCGCCCCGGGCGCGACCTCGTGCGTACCGGCGGGCCAGGGGCACTGCCGGC
25 TTGTGGGGGAGTCGACGACCCGGTGGGAGGGGTCTTTGCCAACGCCGCCGACGCCCCGCCAG
GTCGGCTACCGTCGTGGCTGCGTCCCCGGCTGGGCGACGAGCATGCGCGAGAGAGGACGCCC
GATGCAGGTCGCGATCGCACTTTTTCCGGGCAACACCGCGCTTGA

>Rv0106 - TB.seq 124372:125565 MW:43701

30 >emb|AL123456|MTBH37RV:124372-125568, Rv0106 SEQ ID NO:14
ATGCGTACTCCGGTGATATTGGTGGCAGGTCAGGATCACACCGACGAGGTGACGGGCGCCTTG
TTGCGCCGGACCGGAACGGTGGTCGTGGAGCACCGGTTTGACGGCCATGTGGTGCGACGGAT
GACTGCCACGCTGAGCCGTGGCGAATTGATCACACGGAGGACGCTTTGGAGTTGCCCCACGG
CTGTGTGTCGTGCACAATCCGCGACGACCTGCTGGTGCTGTTACGCAGACTGCACCGCCGAGA
35 CAATGTCGGCCGGATCGTCGTGCACCTGGCGCCGTGGCTGGAGCCCCAGCCCATCTGCTGGG
CGATCGACCACGTGCGGGTTTGCGTCGGACACGGATACCCAGACGGACCAGCCGCCCTCGAC
GTGCGGGTCGCGGCCGTGGTGACCTGTGTGGACTGCGTAAGGTGGCTGCCGCAGTCACTCGG

CGAGGACGAACTGCCCCGACGGGCGCACGGTGGCCCAAGTGACGGTCGGTCAGGCCGAGTTCCG
CCGACCTTCTGGTGCTGACCCACCCGGAACCGGTGCGCGTGGCGGTTCTGCGCCGACTGGCC
CCTCGAGCGCGAATCACCGGCGGCGTGCACCGCGTCGAGCTGGCGCTGGCGCATCTGGACGA
CAACTCACGGAGGGGTCGTACCGATACCCCGCACACGCCATTGCTGGCGGGCCTGCCTCCGTT
5 GGCAGCCGACGGTGAGGTTGCGATCGTGGAATTCAGTGCCCCGCCGCCGTTTCACCCGCAACG
TCTGCATGCCGCGGTTGACCTGCTGCTCGATGGCGTGGTTTCGCACTCGAGGTCGGCTGTGGCT
GGCCAACCGGCCGGATCAGGTCATGTGGCTCGAATCAGCCGGTGGCGGTCTGCGGGTCGCAT
CGGCCGGAAGTGTTGGCGGCGATGGCGGCCTCGGAGGTGGCCTATGTCGACCTGGAGCGG
CGGTTGTTGCCGACCTGATGTGGGTCTACCCGTTTCGGAGACCGGCACACCGCGATGACGGTA
10 CTGGTATGCGGCGCCGATCCGACCGACATCGTCAATGCCCTGAACGCGGCGCTGCTCAGCGAC
GACGAAATGGCATCTCCGCAACGCTGGCAGTCCTACGTCGACCCTTTCGGCGACTGGCATGAC
GACCCGTGCCACGAAATGCCCGATGCGGCTGGGGAATTCTCGGCACACCGCAACTCAGGAGAA
TCTCGATGA

15 >Rv0125 - TB.seq 151146:152210 MW:34927

>emb|AL123456|MTBH37RV:151146-152213, pepA SEQ ID NO:15

ATGAGCAATTCGCGCCGCCGCTCACTCAGGTGGTCATGGTTGCTGAGCGTGCTGGCTGCCGTC
GGGCTGGGCCTGGCCACGGCGCCGCCAGGCGGCCCGCCGGCCTTGTCGCAGGACCGGT
TCGCCGACTTCCCCGCGCTGCCCCCTGACCCGTCCGCGATGGTCGCCCAAGTGGGGCCACAG
20 GTGGTCAACATCAACACCAAACCTGGGCTACAACAACGCCGTGGGCGCCGGGACCGGCATCGTC
ATCGATCCCAACGGTGTCTGTGCTGACCAACAACACGTGATCGCGGGCGCCACCGACATCAAT
GCGTTCAGCGTCGGCTCCGGCCAAACCTACGGCGTCGATGTGGTCGGGTATGACCGCACCCAG
GATGTCGCGGTGCTGCAGCTGCGCGGTGCCGGTGGCCTGCCGTGGCGGGCGATCGGTGGCG
GCGTCGCGGTTGGTGAGCCCGTCTGTCGCGATGGGCAACAGCGGTGGGCAGGGCGGAACGCC
25 CCGTGCGGTGCCTGGCAGGGTGGTCGCGCTCGGCCAAACCGTGCAGGCGTCGGATTGCTGA
CCGGTGCCGAAGAGACATTGAACGGGTTGATCCAGTTCGATGCCGCGATCCAGCCCGGTGATT
CGGGCGGGCCCGTCTGTCACCGGCCTAGGACAGGTGGTCGGTATGAACACGGCCGCGTCCGAT
AACTTCAGCTGTCCCAGGGTGGGCAGGGATTGCCATTCCGATCGGBCAGGCGATGGCGATC
GCGGGCCAGATCCGATCGGGTGGGGGGTCACCACCGTTTCATATCGGGCCTACCGCCTTCCTC
30 GGCTTGGGTGTTGTCGACAACAACGGCAACGGCGCACGAGTCCAACGCGTGGTCGGGAGCGC
TCCGGCGGCAAGTCTCGGCATCTCCACCGGCGACGTGATCACCGCGGTGACGGCGCTCCGAT
CAACTCGGCCACCGCGATGGCGGACGCGCTTAACGGGCATCATCCCGGTGACGTCATCTCGGT
GACCTGGCAAACCAAGTCGGGCGGCACGCGTACAGGGAACGTGACATTGGCCGAGGGACCCC
CGGCCTGA

35

>Rv0350 dnaK 70 kD heat shock protein, chromosome replication TB.seq 419833:421707

MW:66832 SEQ ID NO:16

>emb|AL123456|MTBH37RV:419833-421710, dnaK

ATGGCTCGTGCGGTGCGGATCGACCTCGGGACCACCAACTCCGTCTCGGTTCTGGAAGGT
GGCGACCCGGTCTGCTCGTCCAACTCCGAGGGCTCCAGGACCACCCCGTCAATTGTCGCGTTC
GCCCCGAACGGTGAGGTGCTGGTCCGCCAGCCCCGCAAGAACCAGGCAGTGACCAACGTCGA
5 TCGCACCGTGCCTCGGTCAAGCGACACATGGGCAGCGACTGGTCCATAGAGATTGACGGCAA
GAAATACACCGCGCCGGAGATCAGCGCCCGCATTCTGATGAAGCTGAAGCGCGACGCCGAGGC
CTACCTCGGTGAGGACATTACCGACGCGGTTATCACGACGCCCGCCTACTTCAATGACGCCAG
CGTCAGGCCACCAAGGACGCCCGCCAGATCGCCGGCCTCAACGTGCTGCGGATCGTCAACGA
GCCGACCGCGGCCGCGCTGGCCTACGGCCTCGACAAGGGCGAGAAGGAGCAGCGAATCCTGG
10 TCTTCGACTTGGGTGGTGGCACTTTCGACGTTTCCCTGCTGGAGATCGGCGAGGGTGTGGTTGA
GGTCCGTGCCACTTCGGGTGACAACCACCTCGGCGGCGACGACTGGGACCAGCGGGTCTGTCG
ATTGGCTGGTGGACAAGTTCAAGGGCACCAGCGGCATCGATCTGACCAAGGACAAGATGGCGA
TGCAGCGGCTGCGGGAAGCCGCCGAGAAGGCAAAGATCGAGCTGAGTTGAGTCAGTCCACCT
CGATCAACCTGCCCTACATCACCGTCGACGCCGACAAGAACCCGTTGTTCTTAGACGAGCAGCT
15 GACCCGCGCGGAGTTCCAACGGATCACTCAGGACCTGCTGGACCGCACTCGAAGCCGTTCCA
GTCGGTGATCGCTGACACCGGCATTTCCGTGTCGGAGATCGATCACGTTGTGCTCGTGGGTGG
TTCGACCCGGATGCCCGCGGTGACCGATCTGGTCAAGGAACTCACCGGCGGCAAGGAACCCAA
CAAGGGCGTCAACCCCGATGAGGTTGTGCGGTTGGGAGCCGCTCTGCAGGCCGGCGTCTCA
AGGGCGAGGTGAAAGACGTTCTGCTGCTTGATGTTACCCCGCTGAGCCTGGGTATCGAGACCA
20 AGGGCGGGGTGATGACCAGGCTCATCGAGCGCAACACCACGATCCCCACCAAGCGGTGCGAG
ACTTTCACCACCGCCGACGACAACCAACCGTCGGTGCAGATCCAGGTCTATCAGGGGGAGCGT
GAGATCGCCGCGCACACAAGTTGCTCGGGTCCTTCGAGCTGACCGGCATCCCGCCGGCGCC
GCGGGGGATTCCGCAGATCGAGGTCACTTCGACATCGACGCCAACGGCATTGTGCACGTAC
CGCCAAGGACAAGGGCACCAGGCAAGGAGAACACGATCCGAATCCAGGAAGGCTCGGGCCTGT
25 CCAAGGAAGACATTGACCGCATGATCAAGGACGCCGAAGCGCACGCCGAGGAGGATCGCAAGC
GTCGCGAGGAGGCCGATGTTTCGTAATCAAGCCGAGACATTGGTCTACCAGACGGAGAAGTTG
TCAAAGAACAGCGTGAGGCCGAGGGTGGTTGGAAGGTACCTGAAGACACGCTGAACAAGGTTG
ATGCCGCGGTGGCGGAAGCGAAGCGGCACCTTGGCGGATCGGATATTTCCGCCATCAAGTCG
GCGATGGAGAAGCTGGGCCAGGAGTCGCAGGCTCTGGGGCAAGCGATCTACGAAGCAGCTCA
30 GGCTGCGTCACAGGCCACTGGCGCTGCCACCCCGGCGGCGAGCCGGGCGGTGCCACCCC
GGCTCGGCTGATGACGTTGTGGACGCGGAGGTGGTCGACGACGCCGGGAGGCCAAGTGA

>Rv0351 grpE stimulates DnaK ATPase activity TB.seq 421707:422411 MW:24501

>emb|AL123456|MTBH37RV:421707-422414, grpE SEQ ID NO:17

35 GTGACGGACGGAATCAAAAGCCGGATGGCAATTCGGGCGAACAGGTAACCGTCACTGACAAG
CGGCGGATCGATCCCGAGACGGGTGAAGTGCGGCACGTCCCTCCCGGCGACATGCCGGGAGG
GACGGCTGCGGCCGATGCGGCGCACACCGAAGACAAGGTCGCCGAGCTGACCGCCGATCTGC

AACGCGTGCAGGCCGACTTCGCCAACTACCGTAAGCGGGCGTTGCGCGATCAGCAGGCGGCC
GCTGACCGAGCCAAGGCCAGCGTTGTCAGCCAATTGCTGGGTGTAAGGACGATCTCGAGCGG
GCGCGCAAGCACGGCGATTTGGAGTCGGGTCCACTGAAGTCGGTCGCCGACAAGCTAGACAGC
GCGTTGACCGGGCTGGGTCTGGTGGCGTTCCGGTGCCGAGGGCGAGGATTTGACCCCGTGCT
5 GCACGAAGCGGTGCAACACGAGGGCGACGGCGGGCAGGGGTCCAAGCCGTAATCGGCACC
GTCATGCGGCAGGGCTACCAACTGGGTGAGCAGGTGCTGCGGCACGCCTTGGTGGCGTCGT
CGACACGGTGCTCGTCGACGCGGCCGAAGTGGAGTCAGTCGACGACGGCACTGCGGTGCGCAG
ATACCGCCGAAAACGATCAAGCTGACCAGGGCAATAGCGCCGACACCTCGGGCGAACAGGCAG
AATCAGAACCGTCGGGCAGTTAA

10

>Rv0352 dnaJ acts with GrpE to stimulate DnaK ATPase TB.seq 422450:423634 MW:41346

>emb|AL123456|MTBH37RV:422450-423637, dnaJ SEQ ID NO:18

ATGGCCCAAAGGGAATGGGTGCAAAAAGACTTCTACCAGGAGCTGGGCGTCTCCTCTGATGCC
AGTCCTGAAGAGATCAAACGTGCCTATCGGAAGTTGGCGCGCGACCTGCATCCGGACGCGAAC
15 CCGGGCAACCCGGCCGCCGCGCAACGGTTCAAGGCGTTTCGGAGGCGCATAACGTGCTGTC
GGATCCGGCCAAGCGCAAGGAGTACGACGAAACCCGCCCGCTTTCGCCGGCGGGCGGGTTCCG
GCGGCCGTGCGTTTCGACAGCGGCTTTGGGGGCGGGTTTCGGCGGTTTCGGGGTGGTGGAGAC
GGCGCCGAGTTCAACCTCAACGACTTGTTCGACGCCGCCAGCCGAACCGGCGGTACCACCATC
GGTGACTTGTTCGGTGGCTTGTTCGGACGCGGTGGCAGCGCCCGTCCCAGCCGCCCGCGACG
20 CGGCAACGACCTGGAGACCGAGACCGAGTTGGATTTCTGAGGAGCGCCAAAGGGCGTGGCGA
TGCCGCTGCGATTAACAGCCCGGCGCCGTGCACCAACTGCCATGGCAGCGGGGCCCGGCCA
GGCACCAGCCCAAAGGTGTGTCCCACTTGAACGGGTGCGGCGTGATCAACCGCAATCAGGGC
GCGTTCCGGTTCTCCGAGCCGTGCACCGACTGCCGAGGTAGCGGCTCGATCATCGAGCACCCC
TGCGAGGAGTGCAAAGGCACCGCGTGACCACCCGACCCGAACCATCAACGTGCGGATCCC
25 GCCCGGTGTCGAGGATGGGCAGCGCATCCGGCTAGCCGGTCAGGGCGAGGCCGGGTTGCGC
GGCGCTCCCTCGGGGATCTCTACGTGACGGTGCATGTGCGGCCCGACAAGATCTTCGGCCCG
GACGGCGACGACCTCACCGTCACCGTTCCGGTCAGCTTACCGAATTGGCTTTGGGCTCGACG
CTGTGCGTGCCTACCCTGGACGGCACGGTCGGGGTCCGGGTGCCCAAAGGCACCGCTGACGG
CCGCATTCTGCGTGTGCGCGGACGCGGTGTGCCAAAGCGCAGTGGGGGTAGCGGCGACCTAC
30 TTGTCACCGTGAAGGTGGCCGTGCCGCCCAATTTGGCAGGCGCCGCTCAGGAAGCTCTGGAAG
CCTATGCGGCGGCGGAGCGGTCCAGTGTTTCAACCCGCGGGCCGGATGGGCAGGTAATCGC
TGA

35

>Rv0363c fba fructose bisphosphate aldolase TB.seq 441266:442297 MW:36545

>emb|AL123456|MTBH37RV:c442297-441263, fba SEQ ID NO:19

ATGCCTATCGCAACGCCCGAGGTCTACGCGGAGATGCTCGGTGAGGCCAAACAAAACCTCGTAC
GCTTTCCCGGCTATCAACTGCACCTCCTCGGAAACCGTCAACGCCGCGATCAAAGGTTTCGCCG

ACGCCGGCAGTGACGGAATCATCCAGTTCTCGACCGGTGGCGCAGAATTCGGCTCCGGCCTCG
GGGTCAAAGACATGGTGACCGGTGCGGTGCGCTTGGCGGAGTTCACCCACGTTATCGCGGCCA
AGTACCCGGTCAACGTGGCGCTGCACACCGACCACTGCCCCAAGGACAAGTTGGACAGCTATG
TCCGGCCCTTGCTGGCGATCTCGGCGCAACGCGTGAGCAAAGGTGGCAATCCTTTGTTCCAGT
5 CGCACATGTGGGACGGCTCGGCAGTGCCAATCGATGAGAACCTGGCCATCGCCCAGGAGCTGC
TCAAGGCGGCGGCGGCCGCAAGATCATTCTGGAGATCGAGATCGGCGTCGTGCGCGGCGAA
GAGGACGGCGTGGCGAACGAGATCAACGAGAAGCTGTACACCAGCCCGGAGGACTTCGAGAAA
ACCATCGAGGCGCTGGGCGCGGTGAGCACGGCAAATACCTGCTGGCCGCGACGTTTCGGCAA
CGTGATGGCGTCTACAAGCCCGGCAACGTCAAGCTTCGCCCCGACATCCTTGCGCAAGGGCA
10 ACAGGTGGCGGCGGCCAAGCTCGGACTGCCGGCCGACGCCAAGCCGTTTCTGACTTCGTGTTCC
ACGGCGGCTCGGGTTCGCTTAAGTCGGAGATCGAGGAGGCGCTGCGCTACGGCGTGGTGAAG
ATGAACGTCGACACCGACACCCAGTACGCGTTCACCCGCCCGATCGCCGGTCACATGTTACCC
AACTACGACGGAGTGCTCAAGGTCGATGGCGAGGTGGGTGTCAAGAAGGTCTACGACCCGCGC
AGCTACCTCAAGAAGGCCGAAGCTTCGATGAGCCAGCGGGTCGTTACGGCGTGCAATGACCTG
15 CACTGCGCGGAAAGTCCCTAACCCACTAA

>Rv0405 pks6 TB.seq 485729:489934 MW:147615 >emb|AL123456|MTBH37RV:485729-489937,
pks6 SEQ ID NO:20

ATGACAGACGGTTCGGTCACTGCGGATAAGCTTCAAAAATGGTTTCGAGAGTACTTGTCCACGC
20 ATATCGAGTGTTCATCAAATGAGGTCAGCCTAGACGTTCCGATTAGAGATTTAGGTTTGAATCG
ATTGATGTCTTAGCGATTCCCGGCGACCTCGGTGACAGATTTGGGTTTTGTATTCCCGATTGGC
CGTTTGGGATAATCCTAGCGCTAATGATTTGATTGATAGTCTGTTGAACCAGCGTAGTGCTGACT
CGTTAAGAGAGAGTCATGGACACGCCGACAGGAACACGCAGGGTCGGGGCAGCATAAACGAGC
CGGTTGCGGTTCATCGAGTGGGCTGTCGATTCCGGGAGATATTGACGGCCCGGAACGGCTAT
25 GGGACTTTCTGACCGAGAAGAAGTGTGCGATAACAGCGTATCCAGATCGTGGGTTACGAATGC
TGGAACTTTCGCGGAGTCCGGAGGCTTTTTAAAGGATGTCGCGGGTTTCGATAATAGATTTTTTG
ATATCCCGCCGGACGAGGCTCTGCGAATGGATCCGCAACAACGGTTGTTACTGGAGGTCTCTTG
GGAAGCGTTAGAGCATGCAGGAATTATCCTGAGTCATTAAGACTTTCACGTACGGGCGTATTC
GTTGGGGTGTCTCAACTGACTACGTCCGGCTTGTCGTCAGCTAGCGCTCAGCAAAAGTCTACTA
30 TTTGGGATAACACCGCGGTTCTTCGAGTATTATTGCCAATAGAATCTCATACTTTCTCGATATTC
AGGGTCCGTCCATTGTCATTGACACGGCATGCTCGTCATCCCTGGTCGCCGTGCATCTAGCCTG
TCGAAGTCTCAGTACCTGGGACTGCGATATCGCACTTGTCGGTGGGACGAATGTTCTTATTTAC
CAGAACCATGGGGTGGGTTTAGGGAAGCGGGCATCTTGTCGACAGAGGCTGCTGTCACGCGT
TCGATAAATCCGCCGACGGGATGGTACGCGGTGAGGGATGCGGAGTTATCGTGCTGCAGCGCC
35 TCAGTGATGCACGCCTTGAGGGCCGGCGGATATTAGCGATTCTGACGGGTTACGCGGTCAATC
AGGACGGTAAGTCCAACGGTATTATGGCGCCAAATCCTAGTGCGCAAATTGGTGTCTTGAAAAT
GCATGCAAGAGCGCTCGCGTCGATCCGCTGGAAATCGGCTACGTGAGGCCACGGGACCGG

AACGTCGTTAGGGGATAGGATCGAGGCGCACGCCTTAGGCATGGTCTTTGGTCGCAAGAGACC
GGGATCTGGGCCCCTGATGATCGGGAGCATCAAGCCGAATATCGGCCATCTGGAAGGTGCGGC
TGGCATCGCCGGATTGATCAAGGCGGTGTTGATGGTTGAGCGTGGCTCGCTGCTTCCGAGCGG
GGGGTTTACGGAGCCAAATCCAGCTATCCATTACGGAATTGGGCCTGAGAGTTGTAGACGAA
5 CTTCAGGAGTGCGCGGTGGTGGCGGGTCGGCCGCGCGGGCTGGGGTGTCATCGTTCGGCTT
TGGCGGCACCAATGCGCATGTGATTGTCGAGGAAGCTGGTTCGGTTGGGCGGACACGGTTTC
GGGCCGCGCGGATGTTGGCGGTTCCGGTGGTGGGGTGGTGGCGTGGGTGATTCGGGGAAGA
CGGCTTCGGCGTTGGCTGCTCAGGCGGGTCGGTTGGGCGGTATGTGCGGGCTCGGCCGCGC
CTTGATGTTGTTGATGTGGGGTATTCGTTGGTGAGCACGCGGTGGTGTTCATCATCGGGCGG
10 TGGTGGTTCGGCCAGACTCGCGATGAGTTGCTGGCTGGGTGGCTGGGGTGGTTGCTGGTCGG
CCGGAGGCTGGGGTGGTCTGCGGTGTTGGCAAGCCGGCGGGCAAGACGGCTTTTGTGTTTGC
CGGTCAGGGCTCGCAGTGGCTGGGTATGGGTAGCGAGCTTTATGCTGCCTACCCGGTTTTTCGC
CGAGGCCCTCGATGCTGTGGTGGACGAGTTGGACCGGCACCTGCGGTATCCGCTGCGCGATGT
GATCTGGGGGCACGACCAAGATCTGTTGAATACCACCGAATTCGCCCAGCCGGCGCTGTTTGC
15 GGTGGAGGTGGCGCTGTATCGGCTGCTCATGTCTGTTGGGGGGTGGCGCCGGTTTGGTGCTGG
GTCATTCGGTGGGCGAGTTGGCCGCGGCGCACGTCGCCGGGGCGCTGTGTTGCCGGATGCG
GCGATGCTGGTGGCCGCGCGTGGACGGTTGATGCAGGCGTTGCCCGCCGGCGGCGCCATGTT
TGCGGTGCAGGCCCGTGAAGACGAGGTAGCGCCGATGCTGGGGCACGATGTGAGCATCGCGG
CGGTCAATGGTCCGGCTTCGGTGGTGATCTCTGGTGCCACGATGCGGTGAGCGCGATCGCTG
20 ATCGGCTGCGCGGCCAGGGCCGTCGGGTCCACCGGTTGGCGGTCTCGCATGCCTTCACTCG
GCGTTGATGGAGCCGATGATCGCTGAGTTCACAGCCGTTGCGGCCGAACTGTCTGTGGGCTTG
CCCACGATCCCGGTCATTTCCAATGTGACCGGGCAGTTGGTGGCCGACGACTTCGCCTCAGCT
GATTACTGGGCCCGGCATATCCGGGCGGTGGTGCGGTTTGGCGACAGTGTTCTAGTGCCAC
TGCGCCGGTGCCAGTCGTTTCATCGAAGTCGGGCCCGGTGGCGGCTTGACGTCGTTGATCGAG
25 GCATCGCTGGCCGACGCGCAGATCGTGTGGTGCCACGCTGCGCAAAGATCGGCCCGAACC
GGTCAGTGTGATGACGGCGCGCGGCCAGGGCTTCGTCTCGGGGATGGGCCTGGATTGGGCCT
CGGTGTTTTCCGGGTACCGGCCCAAGCGGGTGGAGTTGCCGACGTATGCCTTCAGCATCAAA
AGTTCTGGCTCGCACCAGCCCCATCGGTCAGCGACCCACCGCCCGCGGCCAGATCGGGGCT
AGCGATGGTGGTGCTGAACTCTTGGCGTCCTCCGGGTTTGCCGCCCGGCTGGCCGGTCCGGTCG
30 GCCGACGAGCAACTCGCCGACGCGATCGAGGTGGTATGTGAGCATGCCGACGCGGTGCTGGG
GCGCGACGGCGCTGCCGGACTCGACGCTGGCCAGGCGTTTGCCGATTCGGGATTTAATTCCTT
GAGTGCCGTGGAGCTACGTAACCGCTTAACAGCCGTCACCGCAGTAACGCTGCCGGCCACCGC
GATCTTCGATCACCCACCCCGACCGAAGTACCCAGTATCTGATCACCCAAATAGACGGTCAC
GGCAGCTCCGCCGCCGACGCGGCAACCCGGCGGAGCGAATCGATGCGCTCACCGATCTTTTT
35 CTACAAGCTTGCGATGCGGGTCGGGATGCCGATGGTTGGAAGATGGTCGCCCTGGCGTCGAAT
ACGCGCGAGCGCATGAGCTACCGGTTTCGGAACAACGTATCGAAGAACGTGCGACTGCTGGCA
GATGGTATCTCCGATGTGGTTGTAATTGTATCCCAACTCTAACTGTGCTATCGGATCAGCGTGA

ATATCGAGATATTGCGAATGCGATGACAGGCCGCCATTTCGGTTTATTCGCTTACGCTTCCCGGG
TTCGATTTCGTCTGATGCACTGCCGCAAAACGCGGATATGATTGTTGAAACCGTATCTAACGCAAT
TATTGATGTGGTAGGCGGCAGCTGCCGTTTTGTGCTGTCGGGCTATTCATCGGGTGGGGTGTG
GCCTATGCCCTCTGCTCCCATCTGTCCGTCAAGCACCAGCGGAATCCCCTCGGAGTCGCACTCA
5 TCGATACATATCTGCCTAGTCAGATCGCCAATCCTTCAATGAATGAAGGGTTCAGCCCCAACGAT
ACTGGGAAGGGCCTTTCCCGTGAAGTAATTCGAGTGGCCAGAATGTTGAATCGGTAACTGCCA
CCCGACTCACCGCGGCAGCCACCTATGCTGCAATCTTCAGGCCTGGGAACCAGGTAGATCAAT
GGCTCCGGTTCTTAACATCGTGGCGAAGGACCGAATAGCTACCGTCGAAAATTTACGCGAAGAA
CGAATCAACCGGTGGCGAACTGCTGCTGCAGAGGCGGCCTATTCTGTAGCCGAAGTACCCGGG
10 GATCATTTTCGGAATGATGAGCACCTCGAGTGAGGCAATAGCTACCGAAATACATGATTGGATTTC
TGGGCTCGTTTCGAGGGCCTCATCGGTAG

>Rv0435c - ATPase of AAA-family TB.seq 522348:524531 MW:75315

>emb|AL123456|MTBH37RV:c524531-522345, Rv0435c SEQ ID NO:21

15 GTGACCCACCCGGACCCGGCCCCGCCAACTACCCCTTACCGCCCGGCTGAACACCTCGGCCGCTC
GACTCACGCCGCGGCGTCTCGGTTGCACCCCAATGCCATTGCTGCCCTTGGCATCCGCGAG
TGGGACGCGGTGTCGCTGACCGGCTCTCGGACAACCGCCGCGGTGCGCGGCCTGGCCGCGGC
AGACACCGCGGTGCGGACGGTGCTGCTCGATGACGTCACACTGTCCAATGCGGGCCTTCGCGA
AGGCACCGAGGTGATCGTCAGCCCGGTACCCGTCTACGGAGCGCGATCGGTGACGCTGAGCG
20 GTTCAACGCTGGCCACCCAGTCGGTGCCGCCGGTCACGCTGCGGCAGGCCCTACTCGGCAAG
GTGATGACCGTCGGTGACGCGGTCTCGTGCTGCCCGCGATCTAGGCCCGGCACATCCACG
TCGGCTGCCAGCCGCGCATTGGCAGCTGCGGTGCGGATCAGTTGGACCTCGGAGCTGCTGACC
GTTACCGGCGTCGACCCCGACGGGCCGGTCAGCGTGACGCCCAACTCGCTGGTCACTTGGGG
CGCTGGGGTCCCGGCCGCAATGGGTACGTCCACGGCCGGGCAAGTGAGCATCTCGAGTCCGG
25 AGATCCAGATCGAAGAGCTCAAGGGCGCCCAGCCGCAGGCTGCCAAGCTCACCGAATGGCTCA
AGCTTGCCCTCGATGAGCCGCACCTACTACAGACCTTGGGCGCCGGCACCAATTTGGGTGTGC
TGGTGTGCGGTCCGGCCGGGGTGGGCAAGGCGACGCTGGTGCGCGCGGTGTGCGACGGCCG
AAGGTTGGTGACACTGGATGGTCCGGAGATTGGAGCTCTGGCCGCCGGAGACCGGGTCAAAGC
CGTGGCCTCGGCAGTGACGGCGGTTCCGCATGAGGGCGGTGTGTTGCTGATACCGATGCCGA
30 CGCCCTGCTGCCAGCCGCCGCCGAGCCGGTAGCCTCGCTGATCCTGTCCGAGCTGCGTACCG
CGGTGGCCACCGCCGGTGTGGTATTGATCGCCACCTCAGCACGGCCCGATCAACTCGATGCCC
GGCTGCGTTCCCCGAGTTGTGCGACCGGGAGCTTGGCCTGCCGCTGCCCGACGCGGCCACC
CGCAAATCGCTGCTGGAGGCGCTGCTGAATCCGGTTCCTACCGGAGACCTCAACCTCGACGAA
ATCGCCTCCCGCACACCGGGTTTCGTCGTGGCCGACCTGGCTGCGCTGGTTCGCGAGGCGGC
35 GCTGCGGGCAGCGTCTCGAGCCAGTGCCGACGGCCGACCACCGATGCTGCACCAAGACGACC
TCCTCGGTGCGTTGACCGTCATCCGGCCGCTGTCCCGCTCGGCCAGCGACGAAGTCACCGTGG
GTGACGTGACGCTCGACGATGTCCGTGACATGGCCGCGGCCAAACAAGCACTGACCGAGGCG

GTGCTGTGGCCGCTGCAGCACCCCGACACCTTCGCTCGGCTAGGTGTGGAACCGCCGCGCGG
GGTGTGCTGTACGGCCCCGCGGCTGCGGCAAGACCTTTGTGGTTCGTGCCCTGGCCAGCAC
CGGACAGTTGAGCGTGCATGCCGTCAAAGGGTCGGAGCTGATGGACAAGTGGGTGGGCTCCTC
GGAGAAGGCAGTCCGCGAGCTATTCCGGCGGGCCCCGCGACTCCGCGCCGTCACTGGTGTTC
5 TCGACGAGCTGGACGCTCTGGCGCCACGGCGCGGTGAGAGCTTCGACTCGGGCGTCTCCGAC
CGGGTGGTGGCCGCGCTGCTGACTGAGCTCGACGGTATTGACCCGCTGCGGGATGTCGTGATG
CTAGGCGCGACCAACCGGCCCCGATCTGATAGACCCGGCGCTGCTGCGCCCCGGGGCGGCTAGA
ACGGCTGGTGTTCGTTGAACCGCCCCGACGCTGCCGCTCGCCGCGAAATCCTGCGCACCGCTGG
CAAGTCGATCCCGCTGAGCTCCGACGTGACCTGGACGAGGTGGCAGCCGGACTCGACGGTTA
10 TAGTGCCGCCGACTGTGTGGCGCTGCTGCGCGAAGCCGCGCTTACCGCGATGCGGCGTTCCAT
CGATGCCGCCAACGTCACCGCCGCCGACCTGGCGACCGCGCGAGAAACCGTGCGCGCGTCGC
TGGATCCGCTGCAGGTGGCGTCGCTGCGTAAGTTCGGCACCAAGGGTGACCTTCGGTCCTAG

>Rv0436c pssA CDP-diacylglycerol-serine o-phosphatidyltransferase TB.seq 524531:525388

15 MW:31219 >emb|AL123456|MTBH37RV:c525388-524528, pssA SEQ ID NO:22

ATGATCGGAAAGCCCCGCGGCAGGCGAGGGGTAAACCTGCAGATACTGCCAGCGCGATGAC
GGTGTCTGTCATTTGCGCGGGACTGACCGCAATCAAGTTTTCGCTCGAGCACCAGCCGAAGGC
CGCGATGGCACTGATCGCCGAGCGGCCATCCTCGACGGGCTCGACGGCCGGGTGGCCCCGCA
TCCTGGATGCCAGTCGCGGATGGGCGCAGAGATCGACTCACTGGCCGACGCGGTGAACTTCG
20 GAGTGACACCCGCGCTGGTGCTTTACGTGTGATGTTGTGCAAGTGGCCGGTCCGTTGGGTGG
TCGTGCTGCTCTACGCGGTGTGCGTGGTATTACGGCTGGCGCGGTACAACGCACTGCAGGACG
ACGGAACCCAGCCCGCCTACGCGCATGAATTCTTCGTGGAATGCCCGCGCCGGCGGGCGCG
GTTTCCATGATCGGCCTGCTAGCCCTCAAAATGCAGTTCGGCGAAGGATGGTGGACCTCGGGCT
GGTTCCTCAGCTTTTGGGTGACGGGAACGTCGATACTCTTGGTCAGCGGGATCCCGATGAAAAA
25 GATGCACGCCGTGTCGGTACCACCCAACCTACGCGGCCGCCCTGCTGGCGGTGCTGGCTATCTG
CGCGGCGGGCCGAGTCCTGGCCCCCTACTTGTGATCTGGGTGATCATCATCGCTACATGTGC
CATATTCCTTTCGCGGTGCGCAGCCAGCGCTGGCTTGCCCAACACCCTGAGGTGTGGGACGAC
AAGCCCAAGCAACGGCGCGCGGTGCGGCGCGCGAGCCGCCGGGCGCATCCCTACCGGCCGT
CGATGGCGCGGCTGGGCCTGCGCAAGCCGGGTCGACGGCTGTGA
30

>Rv0440 groEL 260 kD chaperonin 2 TB.seq 528606:530225 MW:56728

>emb|AL123456|MTBH37RV:528606-530228, groEL2 SEQ ID NO:23

ATGGCCAAGACAATTGCGTACGACGAAGAGGCCCGTCGCGGCCCTCGAGCGGGGCTTGAACGC
CCTCGCCGATGCGGTAAAGGTGACATTGGGCCCCCAAGGGCCGCAACGTCGTCCTGGAAGAA
35 GTGGGGTGCCCCACGATCACCACGATGGTGTGTCCATCGCCAAGGAGATCGAGCTGGAGGA
TCCGTACGAGAAGATCGGCGCCGAGCTGGTCAAAGAGGTAGCCAAGAAGACCGATGACGTGCG
CGGTGACGGCACACGACGGCCACCGTCTGCCCCAGGCGTTGGTTCGCGAGGGCCTGCGCA

ACGTCGCGGCCGGCGCCAACCCGCTCGGTCTCAAACGCGGCATCGAAAAGGCCGTGGAGAAG
GTCACCGAGACCCTGCTCAAGGGCGCCAAGGAGGTGAGACCAAGGAGCAGATTGCGGCCAC
CGCAGCGATTTGCGCGGGTGACCAGTCCATCGGTGACCTGATCGCCGAGGCGATGACAAGGT
GGGCAACGAGGGCGTCATCACCGTCGAGGAGTCCAACACCTTTGGGCTGCAGCTCGAGCTCAC
5 CGAGGGTATGCGGTTGACAAGGGCTACATCTCGGGGTACTTCGTGACCGACCCGGAGCGTCA
GGAGGCGGTCTCTGGAGACCCCTACATCCTGCTGGTCAGCTCCAAGGTGTCCACTGTCAAGGA
TCTGCTGCCGCTGCTCGAGAAGGTATCGGAGCCGGTAAGCCGCTGCTGATCATCGCCGAGGA
CGTCGAGGGCGAGGCGCTGTCCACCCTGGTCGTCAACAAGATCCGCGGCACCTTCAAGTCGGT
GGCGGTCAAGGCTCCCGGCTTCGGCGACCGCCGCAAGGCGATGCTGCAGGATATGGCCATTCT
10 CACCGGTGGTCAGGTGATCAGCGAAGAGGTGCGCCTGACGCTGGAGAACGCCGACCTGTGCG
TGCTAGGCAAGGCCCGCAAGGTGCTGGTCACCAAGGACGAGACCACCATCGTCGAGGGCGCC
GGTGACACCGACGCCATCGCCGACGAGTGGCCAGATCCGCCAGGAGATCGAGAACAGCGA
CTCCGACTACGACCGTGAGAAGCTGCAGGAGCGGCTGGCCAAGCTGGCCGGTGGTGTGCGCG
TGATCAAGGCCGGTGCCGCCACCGAGGTGCAACTCAAGGAGCGCAAGCACCGCATCGAGGAT
15 GCGGTTGCGAATGCCAAGGCCCGCGTCGAGGAGGGCATCGTCGCCGGTGGGGGTGTGACGCT
GTTGCAAGCGGCCCGACCCCTGGACGAGCTGAAGCTCGAAGGCGACGAGGCGACCGGCCCA
ACATCGTGAAGGTGGCGCTGGAGGCCCGCTGAAGCAGATCGCCTTCAACTCCGGGCTGGAGC
CGGGCGTGGTGGCCGAGAAGGTGCGCAACCTGCCGGTGGCCACGGACTGAACGCTCAGACC
GGTGTCTACGAGGATCTGCTCGCTGCCGGCGTTGCTGACCCGGTCAAGGTGACCCGTTCCGGCG
20 CTGCAGAAATGCGGCGTCCATCGCGGGGCTGTTCTGACCACCGAGGCCGTGTTGCCGACAAG
CCGGAAAAGGAGAAGGCTTCCGTTCCCGGTGGCGGCGACATGGGTGGCATGGATTTCTGA

>Rv0482 murB TB.seq 570537:571643 MW:38522

>emb|AL123456|MTBH37RV:570537-571646, murB SEQ ID NO:24

25 ATGAAACGGAGCGGTGTGCGTTGCTCTTTGCCGGTGCGCATATTGCCGAGGCGGTCCCGTTG
GCGCCGCTGACCACTTTGCGTGTGGGCCCGATCGCCGACGTGTATCACTTGACCAAGCGCC
GAACAGGTGGTGGCTGCGCTGCGGCACCTGGATTGCGCGGCCAAGACCGGAGCTGACCGCCC
GCTGGTGTGTTGCTGGTGGCTCCAATTTGGTGATCGCCGAGAACCTGACCGACCTGACCGTGGT
GCGGTTGGCCAATAGCGGCATCACCATCGACGGTAACCTGGTGCGGGCCGAGGCCGGTGGCG
30 TCTTCGATGACGTGGTGGTTAGGGCCATCGAACAGGGTCTGGGCGGACTGGAATGCCTGTCTG
GCATCCCAGGATCGGCCGGGGCGACACCCGTGAGAACGTGGGGGCGTATGGCGCGGAGGT
GTCTGACACCATCACTCGGGTTGCGCTTTTGATCGGTGCACGGGTGAGGTGCGTTGGGTATC
CGCGCGCGACCTGCGCTTCGGCTATCGCACGAGCGTGCTCAAACACGCTGATGGGCTTGCGGT
GCCCACCGTGGTCTTGAGGTGGAGTTTGCGCTGGATCCGTGCGGGCCGAGCGCACCGCTGC
35 GCTACGGCGAGCTGATCGCCGCGCTGAATGCGACCAGCGGCGAGCGCGCCGACCCGCAAGCG
GTCCGCGAAGCGGTGCTGGCCCTGCGGGCACGCAAGGGCATGGTGCTGGACCCGACCGACCA
TGACACCTGGAGCGTGGGATCGTTCTTCACAAACCCGGTGGTCAACCAGGATGTTTACGAACGG

CTGGCCGGTGACGCGGCCACCAGAAAGGACGGTCCGGTCCCGCACTATCCCGCGCCCGACGG
CGTCAAGCTGGCCGCCGGCTGGCTGGTGGAAACGGGCGCGCTTCGGCAAGGGCTATCCGGATG
CCGGCGCCGCCCATGCCGGCTTTCCACCAAACATGCGCTGGCGCTGACAAATCGTGGCGGG
GCCACCGCCGAAGATGTGGTGACGCTGGCGCGCGCCGTGCGCGATGGGGTCCATGATGTGTTT
5 GGTATCACACTAAAACCCGAACCCGTGCTGATCGGCTGCATGTTGTAG

>Rv0483 - TB.seq 571708:573060 MW:47859

>emb|AL123456|MTBH37RV:571708-573063, Rv0483 SEQ ID NO:25

GTGGTCATTCTGTGTGCTGTTTCGCCCGGTATCTTTGATACCCGTGAATAACTCCAGCACCCCCCA
10 GAGTCAGGGGCCGATCAGTCGGCGTCTGGCGTTGACGGCCCTTGGGTTTGGGGTGTGGCACC
GAACGTTCTGGTCGCGTGCGCCGGCAAAGTGACCAAGCTGGCCGAGAAGAGGCCGCCACCGG
CGCCTCGTCTGACTTTCCGGCCTGCCGACTCTGCCGCCGACGTGGTGCCGATCGCGCCGATCA
GCGTCGAGGTCCGTGACGGCTGGTTTCAGCGGGTCGCGCTGACCAATTCGGCAGGCAAGGTC
GTCGCCGGGGCATAACAGCCGGGATCGCACCATCTACACGATCACCGAGCCGCTGGGCTACGAC
15 ACGACCTACACCTGGAGCGGTTCCGGCCGTGCGCCATGACGGCAAGGCGGTTCCGGTGGCGGG
CAAGTTCACCACCGTGGCACCCGTCAAGACGATCAACGCGGGATTCCAGCTCGCCGACGGCCA
GACCGTCGGGATCGCGGCGCCGGTGATTATTAGTTTCGATTACCGATCAGCGACAAGGCCGC
CGTCGAGCGGGCACTAACCGTGACCACCGACCCGCCTGTCGAGGGCGGCTGGGCCTGGCTGC
CCGACGAGGCGCAGGGCGCTCGCGTGCACTGGCGTCCTCGGGAGTACTACCGGGCGGGTACC
20 ACCGTCGACGTCGACGCCAAGCTGTATGGGCTGCCGTTCCGGCGACGGCGCGTACGGCGCGCA
GGATATGTCGTTGCACTTCAGATCGGTGCTCGTCAGGTGGTCAAGGCCGAAGTCTCGTCGCAC
CGCATCCAAGTCGTCACCGATGCCGGCGTCATCATGGACTTCCCGTGACGCTACGGCGAGGCC
GACTTGGCGCGCAACGTACCCGCAACGGCATCCACGTCGTCACCGAGAAATACTCGGACTTC
TACATGTCCAACCCGGCCGCGGTTACAGCCATATCCACGAACGTTGGGCGGTGCGGATTTCC
25 AACAACGGCGAGTTCATCCATGCCAACCCTATGAGCGCCGGTGCCCAGGGCAACAGCAATGTC
ACCAACGGCTGTATCAACCTGTGACGGAGAACGCCGAACAGTACTACCGCAGCGCGGTCTAC
GGTGACCCGGTTGAGGTGACCGGCAGTTCGATCCAGCTGTCCTACGCCGACGGTGACATCTGG
GACTGGGCGGTGGACTGGGACACCTGGGTGTCGATGTCGGCGCTACCGCCACCGGCGGCCAA
ACCGGCGGGCGACGCAAATCCCGGTCACCGCCCCGGTCACGCCGTGCGATGCCCCACCCCGT
30 CCGGCACACCCACGACTACTAACGGACCGGGTGGGTAG

>Rv0489 gpm phosphoglycerate mutase I TB.seq 578424:579170 MW:27217

>emb|AL123456|MTBH37RV:578424-579173, gpm SEQ ID NO:26

ATGGCAAACACTGGCAGCCTGGTGTGCTGCGCCACGGCGAGAGCGACTGGAATGCCCTCAAC
35 CTGTTACCGGCTGGGTGATGTGCGCCTGACGGACAAGGGCCAGGCAGAGGCGGTTGGAAG
CGGCGAGCTGATCGCGGAACACGACCTATTGCCCCAGCTGCTCTACACCTCGTTGCTGCGGCG
CGCGATCACCACCGCGCATCTGGCGTTGGACAGCGCCGATCGGCTCTGGATTCCCGTGCGGCG

TAGCTGGCGGCTCAACGAACGCCACTACGGCGCGCTGCAGGGTTTGGACAAGGCCGAGACCAA
GGCCCGCTATGGCGAAGAGCAGTTCATGGCCTGGCGGCGCAGCTATGACACGCCGCCGCCGC
CGATCGAGCGGGGCAGTCAGTTCAGCCAGGACGCCACCCTCGTTACGCCGACATCGGCGGT
GGCCCGCTCACCGAATGTCTGGCTGACGTGGTCGCCCGGTTTTTGGCATATTTACCCGACGTCA
5 TCGTTGGCGACTTGCAGGGTCGGCAAGACGGTGCTGATCGTTGCCACGGCAACTCGTTGCGCG
CGCTGGTCAAGCACCTGGACCAGATGTCTGACGACGAAATCGTCGGACTGAACATCCCGACCG
GAATTCGGCTGCGCTACGACCTGGATTCCGCGATGAGGCCGCTGGTGCGCGGTGGTACGTATC
TGGACCCGGAGGCGGCAGCCGCCGCCGCCGCCGGTGGCCGCCAGGGCCGCCGGGTAA

10 >Rv0490 senX 3sensor histidine kinase TB.seq 579347:580576 MW:44794
>emb|AL123456|MTBH37RV:579347-580579, senX3 SEQ ID NO:27
GTGACTGTGTTCTCGGCGCTGTTGCTGGCCGGGGTTTTGTCCGCGCTGGCACTGGCCGTCGGT
GGTGCTGTTGGAATGCGGCTGACGTGCGGGTCTGCGAACAGCGCCAACGGGTGGCCACGGA
GTGGTCGGGAATCACGGTTTCGCAGATGTTGCAATGCATTGTCACGCTGATGCCGCTGGGCGC
15 CGCGGTGGTGACACCCATCGCGACGTTGTCTACCTCAACGAACGGGCCAAAGAGCTAGGTCT
GGTGCGCGACCGCCAGCTCGATGATCAGGCCTGGCGGGCCGCCCGGAGGCGCTGGGTGGT
GAAGACGTCGAGTTCGACCTGTGCCGCGCAAGCGGTGCGCCACGGGTGATCCGGGCTATC
AGTGCATGGGCATGCCCGGTTGCTGAGCGAGGAAGACCGCCGGTTCCCGTGGTGTTCGTGCA
CGACCAGTCGGATTATGCGCGGATGGAGGCGGCTAGGCGTGACTTCGTGGCCAACGTCAGTCA
20 CGAGCTCAAGACGCCCGTCGGTGCCATGGCTCTACTCGCCGAGGCGCTGCTGGCGTCGGCCG
ACGACTCCGAAACCGTTTCGGCGGTTCCGCCGAGAAGGTGCTCATTGAGGCCAACCGGCTCGGTG
ACATGGTCGCCGAGTTGATCGAGCTATCCCGGCTACAGGGCGCCGAGCGGCTACCCAATATGA
CCGACGTCGACGTCGATACGATTGTGTGCGGAAGCGATTCACGCCATAAGGTGGCGGCCGACA
ACGCCGACATCGAAGTCCGCACCGACGCGCCAGCAATCTGCGGGTGTGGGCGACCAAACCTC
25 TGCTGGTTACCGCACTGGCAAACCTGGTTTCCAATGCGATTGCCTATTCGCCGCGCGGGTCGCT
GGTGTGATCAGCCGTCGCCGTGCGGGTGCCAACATCGAGATCGCCGTACCGACCGGGGCA
TCGGCATCGCGCCGGAAGACCAGGAGCGGGTCTTCGAACGGTTCTTCGGGGGGACAAGGCCG
CGCTCGCGTGCCACCGGAGGCAGCGGACTCGGGTTGGCCATCGTCAAACACGTCGCGGCTAAT
CACGACGGCACCATCCGCGTGTGGAGCAAACCGGGAACCGGGTCAACGTTACCTTGGCTCTT
30 CCGGCGTTGATCGAGGCCTATCACGACGACGAGCGACCCGAGCAGGCGCGAGAGCCCGAACT
GCGGTCAAACAGGTCAACAGAGAGGAAGAGCTGAGCCGATGA

>Rv0500 proC pyrroline-5-carboxylate reductase TB.seq 590081:590965 MW:30172

>emb|AL123456|MTBH37RV:590081-590968, proC SEQ ID NO:28

35 ATGCTTTTCGGCATGGCAAGGATCGCGATTATCGGCGGCGGCAGCATCGGTGAGGCATTGCTG
TCGGGTCTGCTGCGGGCGGGCCGGCAGGTCAAAGACCTGGTAGTGGCCGAGCGGATGCCCGA
TCGCGCCAACCTACCTGGCGCAGACCTATTCGGTGTTGGTGACGTCGGCGGCCGACGCGGTGGA

GAACGCGACGTTGTCGTCGTCGCGGTCAAACCAGCCGACGTCGAGCCGGTGATCGCGGATCT
GGCGAACGCGACTGCGGCGGCCGAAAACGACAGTGCTGAGCAGGTGTTGTCACCGTGGTAG
CGGGCATCACGATCGCGTATTTTGAATCCAAGCTACCGGCTGGGACGCCAGTGGTGCGTGCGA
TGCCGAACGCGGCGGCATTGGTGGGAGCGGGGGTTACAGCGCTGGCCAAAGGCCGCTTTGTC
5 ACCCCGCAACAGCTTGAGGAGGTCTCGGCCTTGTTGACGCGGTGCGCGGCGTGTGACCGTT
CCGGAATCGCAGTTGGACGCGGTGACCGCGGTGTCCGGCTCGGGTCCGGCCTATTTCTTTCTG
CTGGTCGAGGCCCTGGTGGATGCCGGAGTCGGGGTGGGCTTGAGCCGTCAGGTGGCCACCGA
TCTCGCCGCGCAGACAATGGCTGGCTCAGCGCGGATGCTGCTGGAGCGGATGGAGCAAGACC
AGGGTGGCGCCAATGGCGAGCTGATGGGGTGC CGCTGGACCTTACCGCATCACGGCTGCGC
10 GCCGCGGTTACCTCGCCGGGCGGTACGACCGCCGCTGCGCTGCGGGAACCTCGAACGCGGCG
GGTTTCGGATGGCTGTGACGCGGCGGTTCAAGCCGCCAAAAGCCGCTCTGAGCAGCTCAGAA
TTACACCGGAATGA

>Rv0528 - TB.seq 618303:619889 MW:57132

>emb|AL123456|MTBH37RV:618303-619892, Rv0528 SEQ ID NO:29

15 ATGTGGCGGTGCTTGACGTCGATGGGCACCGCGCTGGTGCTGCTGTTTTGCTCGCGCTGGCT
GCCATACCCGGGGCCCTGCTGCCGACGCTGGCCTCAACGCCGCCAAGGTGGACGACTACCT
GGCCGCGCACCCACTCATCGGTCCGTGGCTGGACGAGCTGCAGGCCTTCGACGTGTTCTCCAG
CTTCTGGTTACCGCCATCTACGTGCTGCTGTTGCTGTCCTCGTCGGCTGTCTGGCCCCGCGG
ACGATCGAGCACGCCCGCAGCCTGCGGGCTACACCGGTGCGCGCCCCGCGCAACCTGGCCCG
20 GCTGCCCAAGCACGCCACGCCCGGCTGGCCGGCGAGCCCGCCGCCCTGGCCGCCACCATCA
CGGGCCGGCTGCGCGGCTGGCGCAGCATCACCCGGCAACAAGGCGACAGCGTGGAAGTCTCC
GCCGAGAAGGGCTACCTGCGCGAGTTCGGCAACCTGGTGTTCACCTTCGCGCTGCTGGGTCTG
CTGGTGGCGGTGGCCGTGCGCAAGCTGTTCCGGCTACGAGGGCAACGTGATCGTGATAGCCGA
CGGCGGACCCGGTTTTTGTTCGGCGTCCCGGCGCGGTTGACTCGTTTCGCGCCGGCAACAC
25 CGTCGACGGCACGTGCTTGACCCGATCTGTGTGCGGGTCAACAACCTCCAAGCGCACTACCT
GCCGTCCGGGCAGGCCACCTCGTTCGCCGCCGACATCGACTATCAGGCCGACCCGGCCACTG
CTGACCTGATCGCCAACAGCTGGCGGCCCTACCGGCTGCAGGTCAATCACCCGCTGCGGGTCC
GCGGCGACCGGGTGTACCTGCAGGGCCACGGCTATGCGCCACCTTACCGTGACGTTCCCG
GACGGGCAGACCCGCACGTGACCGTGCAGTGGCGACCCGACAACCCGCAGACCCTGCTGTC
30 GCGGGGCGTCTGCGCATCGACCCGCCGGCCGGCAGCTACCCCAACCCCGACGAGCGTCGCA
AACACCAGATCGCCATCCAGGGCCTGCTGGCTCCACCGAGCAGCTCGACGGCACCCCTGCTGT
CGTCGCGTTTTCCCGCGCTCAATGCCCGGCGGTGGCCATCGACATCTACCGCGGCGACACCG
GCCTGGACAGCGGGCGGCCCCAGTCGTTGTTACCCCTGGACCACCGGCTGATCGAGCAGGGC
CGGCTGGTCAAGGAAAAGCGGGTCAACCTGCGCGCCGGTCAGCAAGTCCGCATCGACCAAGG
35 CCCGGCGGCCGGCACGGTGGTCCGGTTCGACGGCGCGGTGCCGTTGTCGTAACCTGCAGGTCT
CCCACGACCCCGGCCAGTCCTGGGTGCTGGTCTTCGCAATCAGATGATGGCGGGACTGCTGG
TGTCGCTGCTGGTGCGCAGGCGCCGGGTGTGGGCGCGGATCACGCCGACGACCGCGGGTACG

GTAAACGTCGAGCTGGGCGGCCTGACGCGCACCGACAACCTCCGGGTGGGGCGCCGAGTTCGA
GCGGCTGACCGGGCGGTTGCTGGCGGGTTTTGAGGCGCGTCCCCGGACATGGCCGAAGCGG
CCGCAGGGACCGGAAGGGACGTCGATTGA

5 >Rv0667 rpoB [beta] subunit of RNA polymerase TB.seq 759805:763320 MW:129220
>emb|AL123456|MTBH37RV:759805-763323, rpoB SEQ ID NO:30
TTGGCAGATTCCTCGCCAGAGCAAAACAGCCGCTAGTCCTAGTCCGAGTCGCCCCGAAAGTTCCT
CGAATAACTCCGTACCCGGAGCGCCAAACCGGGTCTCCTTCGCTAAGCTGCGCGAACCACCTG
AGGTTCCGGGACTCCTTGACGTCCAGACCGATTGTTTCGAGTGGCTGATCGGTTCCGCCGCGCT
10 GGC CGCAATCCGCCGCCGAGCGGGGTGATGTCAACCCAGTGGGTGGCCTGGAAGAGGTGCTC
TACGAGCTGTCTCCGATCGAGGACTTCTCCGGGTGATGTGCTTGTCTGTTCTCTGACCCCTCGTT
TCGACGATGTCAAGGCACCCGTCGACGAGTGCAAAGACAAGGACATGACGTACGCGGCTCCAC
TGTTTCGTCACCGCCGAGTTCATCAACAACAACACCGGTGAGATCAAGAGTCAGACGGTGTTTCAT
GGGTGACTTCCCGATGATGACCGAGAAGGGCACGTTTCATCATCAACGGGACCGAGCGTGTGGT
15 GGTCAGCCAGCTGGTGCGGTGCCCCGGGGTGTACTTCGACGAGACCATTGACAAGTCCACCGA
CAAGACGCTGCACAGCGTCAAGGTGATCCCGAGCCGCGGCGCGTGGCTCGAGTTTGACGTGCA
CAAGCGCGACACCGTCGGCGTGCGCATCGACCGCAAACGCCGGCAACCGGTACCGTGCTGC
TCAAGGCGCTGGGCTGGACCAGCGAGCAGATTGTGCGAGCGGTTCCGGTTCTCCGAGATCATGC
GATCGACGCTGGAGAAGGACAACACCGTCGGCACCGACGAGGCGCTGTTGGACATCTACCGCA
20 AGCTGCGTCCGGGCGAGCCCCGACCAAAGAGTCAGCGCAGACGCTGTTGGAAAACCTGTTCT
TCAAGGAGAAGCGCTACGACCTGGCCCCGCTCGGTGCGTATAAGGTCAACAAGAAGCTCGGGC
TGCATGTGCGCGAGCCCATCACGTGTCGACGCTGACCGAAGAAGACGTCGTGGCCACCATCG
AATATCTGGTCCGCTTGACGAGGGTCAGACCACGATGACCGTTCCGGGCGGCGTCGAGGTGC
CGGTGGAAACCGACGACATCGACCACTTCGGCAACCGCCGCTGCGTACGGTCGGCGAGCTG
25 ATCCAAAACAGATCCGGGTCGGCATGTGCGGATGGAGCGGGTGGTCCGGGAGCGGATGAC
CACCCAGGACGTGGAGGCGATCACACCGCAGACGTTGATCAACATCCGGCCGGTGGTCGCCC
CGATCAAGGAGTTCTTCGGCACCGAGCCAGCTGAGCCAATTCATGGACCAGAACAACCCGCTGTC
GGGGTTGACCCACAAGCGCCGACTGTGCGCGCTGGGGCCCCGGGCTGTGCACGTGAGCGTG
CCGGGCTGGAGGTCCGCGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGATCGAAACC
30 CCTGAGGGGCCCAACATCGGTCTGATCGGCTCGCTGTGCGGTGTACGCGCGGGTCAACCCGTT
GGGTTTCATCGAAACGCCGTACCGCAAGGTGGTCGACGGCGTGGTTAGCGACGAGATCGTGAC
CTGACCGCCGACGAGGAGGACCGCCACGTGGTGGCACAGGCCAATTCGCCGATCGATGCGGA
CGGTGCGTTTCGTCGAGCCGCGCGTGCTGGTCCGCCGCAAGGCGGGCGAGGTGGAGTACGTGC
CCTCGTCTGAGGTGGACTACATGGACGTCTCGCCCCGCCAGATGGTGTGCGGTGGCCACCGCGA
35 TGATTCCCTTCTGGAGCACGACGACGCCAACCGTGCCCTCATGGGGGCAAACATGCAGCGCC
AGGCGGTGCCGCTGGTCCGTAGCGAGGCCCGCTGGTGGGCACCGGGATGGAGCTGCGCGC
GGCGATCGACCGCGGCGACGTGCTCGTCGCGGAAGAAAGCGGCGTCATCGAGGAGGTGTGCG

CCGACTACATCACTGTGATGCACGACAACGGCACCCGGCGTACCTACCGGATGCGCAAGTTTG
CCCGGTCCAACCACGGCACTTGCGCCAACCAAGTGCCCCATCGTGACGCGGGCGACCGAGTC
GAGGCCGGTCAGGTGATCGCCGACGGTCCCTGTACTGACGACGGCGAGATGGCGCTGGGCAA
GAACCTGCTGGTGGCCATCATGCCGTGGGAGGGCCACAACCTACGAGGACGCGATCATCCTGTC
5 CAACCGCCTGGTGAAGAGGACGTGCTCACCTCGATCCACATCGAGGAGCATGAGATCGATGC
TCGCGACACCAAGCTGGGTGCGGAGGAGATCACCCGCGACATCCCGAACATCTCCGACGAGGT
GCTCGCCGACCTGGATGAGCGGGGCATCGTGCGCATCGGTGCCGAGGTTCCGCGACGGGGACA
TCCTGGTCGGCAAGGTACCCCCGAAGGTGAGACCGAGCTGACGCCGAGGAGCGGCTGCTG
CGTGCCATCTTCGGTGAGAAGGCCCGCGAGGTGCGCGACACTTCGCTGAAGGTGCCGCACGG
10 CGAATCCGGCAAGGTGATCGGCATTCCGGGTGTTTTCCCGCGAGGACGAGGACGAGTTGCCGGC
CGGTGTCAACGAGCTGGTGCGTGTGTATGTGGCTCAGAAACGCAAGATCTCCGACGGTGACAA
GCTGGCCGGCCGGCACGGCAACAAGGGCGTGATCGGCAAGATCCTGCCGGTTGAGGACATGC
CGTTCCTTGCCGACGGCACCCCGGTGGACATTATTTGAACACCCACGGCGTGCCGCGACGGA
TGAACATCGGCCAGATTTTGAGAGACCCACCTGGGTGGTGTGCCACAGCGGCTGGAAGGTGCG
15 ACGCCGCCAAGGGGGTTCCGGACTGGGCCGCCAGGCTGCCCGACGAACCTGCTCGAGGCGCAG
CCGAACGCCATTGTGTGACGCCGGTGTTGACGGCGCCAGGAGGCCGAGCTGCAGGGCCT
GTTGTCTGTGACGCTGCCCCAACCGCGACGGTGACGTGCTGGTGCACGCCGACGGCAAGGCCA
TGCTCTTCGACGGGCGCAGCGGCGAGCCGTTCCCGTACCCGGTCACGGTTGGCTACATGTACA
TCATGAAGCTGCACCACCTGGTGGACGACAAGATCCACGCCCCGCTCCACCGGGCCGTA CTGA
20 TGATCACCCAGCAGCCGCTGGGCGGTAAGGCGCAGTTCGGTGGCCAGCGGTTCCGGGGAGATG
GAGTGCTGGGCCATGCAGGCCTACGGTGCTGCCTACACCCTGCAGGAGCTGTTGACCATCAAG
TCCGATGACACCGTCGGCCGCGTCAAGGTGTACGAGGCGATCGTCAAGGGTGAGAACATCCCG
GAGCCGGGCATCCCCGAGTCGTTCAAGGTGCTGCTCAAAGAACTGCAGTCGCTGTGCCTCAAC
GTGAGGTGCTATCGAGTGACGGTGCGGCGATCGAACTGCGCGAAGGTGAGGACGAGGACCT
25 GGAGCGGGCCGCGGCCAACCTGGGAATCAATCTGTCCCGCAACGAATCCGCAAGTGTGAGGA
TCTTGCGTAA

>Rv0668 rpoC [beta]' subunit of RNA polymerase TB.seq 763368:767315 MW:146740

>emb|AL123456|MTBH37RV:763368-767318, rpoC SEQ ID NO:31

30 GTGCTCGACGTCAACTTCTTCGATGAACTCCGCATCGGTCTTGCTACCGCGGAGGACATCAGGC
AATGGTCCTATGGCGAGGTCAAAAAGCCGGAGACGATCAACTACCGCACGCTTAAGCCGGAGA
AGGACGGCCTGTTCTGCGAGAAGATCTTCGGGCCGACTCGCGACTGGGAATGCTACTGCGGCA
AGTACAAGCGGGTGCGCTTCAAGGGCATCATCTGCGAGCGCTGCGGCGTCGAGGTGACCCGC
GCCAAGGTGCGTCGTGAGCGGATGGGCCACATCGAGCTTGCCGCGCCCGTCACCCACATCTG
35 GTACTTCAAGGGTGTGCCCTCGCGGCTGGGGTATCTGCTGGACCTGGCCCCGAAGGACCTGGA
GAAGATCATCTACTTCGCTGCCTACGTGATCACCTCGGTGACGAGGAGATGCGCCACAATGAG
CTCTCCACGCTCGAGGCCGAATGGCGGTGGAGCGCAAGGCCGTGGAAGACCAGCGCGACGG

CGAACTAGAGGCCCGGGCGCAAAAGCTGGAGGCCGACCTGGCCGAGCTGGAGGCCGAGGGC
GCCAAGGCCGATGCGCGGCGCAAGGTTGCGGACGGCGGCGAGCGCGAGATGCGCCAGATCC
GTGACCGCGCGCAGCGTGAGCTGGACCGGTTGGAGGACATCTGGAGCACTTTCACCAAGCTGG
CGCCCAAGCAGCTGATCGTCGACGAAAACCTCTACCGCGAACTCGTCGACCGCTACGGCGAGT
5 ACTTCACCGGTGCCATGGGCGCGGAGTCGATCCAGAAGCTGATCGAGAACTTCGACATCGACG
CCGAAGCCGAGTCGCTGCGGGATGTCATCCGAAACGGCAAGGGGCAGAAGAAGCTTCGCGCC
CTCAAGCGGCTGAAGGTGGTTGCGGCGTTCGAACAGTCGGGCAACTCGCCGATGGGCATGGTG
CTCGACGCCGTCCCGGTGATCCCGCCGAGCTGCGCCCGATGGTGACGCTCGACGGCGGCGCG
GTTGCGCCACGTCCGACTTGAACGACCTGTACCGCAGGGTGATCAACCGCAACAACCGGCTGAA
10 AAGGCTGATCGATCTGGGTGCGCCGAAATCATCGTCAACAACGAGAAGCGGATGCTGCAGGA
ATCCGTGGACGCGCTGTTGACAATGGCCGCCGCGGCCGCGCCGTCACCGGGCCGGGCAACC
GTCCGCTCAAGTCGCTTTCGATCTGCTCAAGGGCAAGCAGGGCCGGTTCCGGCAGAACCTGC
TCGGCAAGCGTGTGCGACTACTCGGGCCGGTCCGGTCATCGTGGTCGGCCCGCAGCTCAAGCTGC
ACCAAGTGCAGTCTGCCAAGCTGATGGCGCTGGAGCTGTTCAAGCCGTTCTGTGATGAAGCGGC
15 TGGTGGACCTCAACCATGCGCAGAACATCAAGAGCGCCAAGCGCATGGTGGAGCGCCAGCGCC
CCCAAGTGTGGGATGTGCTCGAAGAGGTCATCGCCGAGCACCCGGTGTGCTGAACCGCGCAC
CCACCCTGCACCGGTTGGGTATCCAGGCCTTCGAGCCAATGCTGGTGGAAGGCAAGGCCATTG
AGCTGCACCCGTTGGTGTGTGAGGCGTTCAATGCCGACTTCGACGGTGACCAGATGGCCGTGC
ACCTGCCTTTGAGCGCCGAAGCGCAGGCCGAGGCTCGCATTTTGATGTTGTCTCCAACAACAT
20 CCTGTGCGCCGBCATCTGGGCGTCCGTTGGCCATGCCGCGGCTGGACATGGTGACCGGGCTGT
ACTACCTGACCACCGAGGTCCCCGGGGACACCGGCGAATACCAGCCGGCCAGCGGGGATCAC
CCGGAGACTGGTGTCTACTCTTCGCCGGCCGAAGCGATCATGGCGGCCGACCGCGGTGTCTTG
AGCGTGCGGGCCAAGATCAAGGTGCGGCTGACCCAGCTGCGGGCCGCGGTGAGATCGAGGC
CGAGCTATTGCGCCACAGCGGCTGGCAGCCGGGCGATGCGTGATGGCCGAGACCACGCTGG
25 GCCGGGTGATGTTCAACGAGCTGCTGCCGCTGGGTTATCCGTTCTGTCACCAAGCAGATGCACAA
GAAGGTGCAGGCCGCCATCATCAACGACCTGGCCGAGCGTTACCCGATGATCGTGGTCGCCCCA
GACCGTCGACAAGCTCAAGGACGCCGGCTTCTACTGGGCCACCCGCAGCGGCGTGACGGTGT
CGATGGCCGACGTGCTGGTGCCGCCGCGCAAGAAGGAGATCCTCGACCACTACGAGGAGCGC
GCGGACAAGGTGAAAAGCAGTTCCAGCGTGGCGCTTTGAACCACGACGAGCGCAACGAGGC
30 GCTGGTGGAGATTTGGAAGGAAGCCACCGACGAGGTCCGTGAGGCGTTGCGGGAGCACTACC
CCGACGACAACCCGATCATCACCATCGTCGACTCCGGCGCCACCGGCAACTTCACCCAGACTC
GAACGCTGGCCGGTATGAAGGGCCTGGTGACCAACCCGAAGGGTGAGTTCATCCCGCGTCCG
GTCAAGTCCTCCTTCCGTGAGGGCCTGACCGTGCTGGAGTACTTCATCAACACCCACGGCGCTC
GAAAGGGCTTGGCGGACACCGCGTTGCGCACCGCCGACTCCGGCTACCTGACCCGACGTCTG
35 GTGGACGTGTCCAGGACGTGATCGTGCGCGAGCACGACTGCCAGACCGAGCGCGGCATCGT
CGTCGAGCTGGCCGAGCGTGACCCGACGGCACGCTGATCCGCGACCCGTACATCGAAACCTC
GGCCTACGCGCGGACCCTGGGCACCGACGCGGTGACGAGGCCGGCAACGTCATCGTCGAGC

GTGGTCAAGACCTGGGCGATCCGGAGATTGACGCTCTGTTGGCTGCTGGTATTACCCAGGTCAA
GGTGCGTTCGGTGCTGACGTGTGCCACCAGCACCGGCGTGTGCGCGACCTGCTACGGGCGTT
CCATGGCCACCGGCAAGCTGGTGCACATCGGTGAAGCCGTCGGCATCGTGGCCGCCAGTCC
ATCGGCGAACCCGGCACCCAGCTGACCATGCGCACCTTCCACCAGGGTGGCGTCGGTGAGGA
5 CATACCCGGTGGTCTGCCCCGGGTGCAGGAGCTGTTGAGGCCCCGGGTACCGCGTGGAAGG
CGCCGATCGCCGACGTACCCGGCCGGGTTTCGGCTCGAGGACGGCGAGCGGTTCTACAAGATC
ACCATCGTTCCTGACGACGGCGGTGAGGAAGTGGTCTACGACAAGATCTCCAAGCGGCAGCGG
CTGCGGGTGTTCAGCACGAAGACGGTTCGAACGGGTGCTCTCCGATGGCGACCACGTGAG
GTGGGCCAGCAGCTGATGGAAGGCTCGGCCGACCCGCATGAGGTGCTGCGGGTGCAGGGCCC
10 CCGCGAGGTGCAGATACACCTGGTTCGCGAGGTCCAGGAGGTCTACCGCGCCCAAGGTGTGTC
GATCCACGACAAGCACATCGAGGTGATCGTTCGCCAGATGCTGCGCCGGGTGACCATCATCGA
CTCGGGCTCGACGGAGTTTTTGCCTGGCTCGCTGATCGACCGCGCGGAGTTCGAGGCAGAGAA
CCGCCGAGTGGTGGCCGAGGGCGGTGAGCCCGCGGCCGGCCGTCCGGTGCTGATGGGCATC
ACGAAGGCGTCGCTGGCCACCGACTCGTGGCTGTGCGCGGCGTCTGTTCCAGGAGACCACTCG
15 CGTGCTGACCGATGCGGCGATCAACTGCCGACGATAAGCTCAACGGTCTGAAGGAAAACGT
GATCATCGGCAAGCTGATCCCGGCCGGTACCGGTATCAACCGCTACCGCAACATCGCGGTGCA
GCCACCGAGGAGGCCCGCGTGTGCGGCGTACACCATCCCGTCGTATGAGGATCAGTACTACAG
CCCGGACTTCGGTGTGCGGCCACCGGTGCTGCCGTCCCGCTGGACGACTACGGCTACAGCGACTA
CCGCTAG

20
>Rv0711 atsa TB.seq 806333:808693 MW:86216
>emb|AL123456|MTBH37RV:806333-808696, atsa SEQ ID NO:32
ATGGCACCCGAGGCCACCGAGGCGTTCAACGGCACCATCGAGCTGGATATTCGTGATTCCGGAG
CCGGATTGGGGCCCATACGCAGCGCCGGTGGCACCGGAGCACTACCAAACATCCTGTATCTG
25 GTCTGGGACGACGTGGGCATCGCGACCTGGGACTGCTTTGGCGGCCTGGTTCGAGATGCCCGC
GATGACGCGCGTCGCCGAGCGTGGCGTGGCACTGTGCAATTTACACCAACCGCACTGTGCTC
GCCGACCCGGGCGTGTGCTGACCGGTGCAACGCCACCAACCGTAGGCATGGCTACCATCG
AAGAGTTACCGACGGGTTCCCCAACTGCAACGGGCGGATCCCGGCTGACACCGCGTTGCTCC
CAGAGGTGCTGGCCGAACATGGCTACAACACCTACTGTGTGGGCAAGTGGCACCTGACGCCAC
30 TCGAAGAATCCAATATGGCGTCGACGAAGCGGCACTGGCCGACCTCGCGTGGGTTTCGAGCGGT
TCTACGGATTCTAGGCGGGGAGACCGACCACTGGTATCCCGACCTGGTATACGACAACCAAC
CAGTGAGTCTCTCCCGGCACACCCGAGGGTGGCTACCACCTGTCAAAAGACATCGCCGACAAGA
CGATCGAGTTCATTCTGTATGCCAAGGTGATCGCGCCCGACAAGCCGTGGTTCAGCTACGTGTG
CCCAGGCGCCGGGCATGCGCCGACCACTCTTCAAGGAATGGGCGGACAGATACGCCGGCC
35 GATTCGACATGGGGTATGAGCGCTATCGCGAGATCGTGTGAAAGGCAAAAGGCGCTAGGGA
TCGTGCCACCCGACACCGAACTGTGCGCCATAAACCTTATCTGGATGTGCCGGGGGCCAAACG
GCGAGACCTGGCCGCTGCAGGACACGGTGTGCGCCGTGGGACTCGCTGAGCGATGAAGAAAAG

AAGCTGTTTTGCCGGATGGCCGAGGTGTTGCCGGCTTTCTGAGCTACACCGACGCCCAGATC
GGACGGATCCTGGACTACCTCGAGGAATCCGGCCAGCTGGACAACACCATCATCGTGGTGATC
TCCGACAACGGCGCCAGCGGCGAGGGCGGACCCAACGGATCGGTCAACGAAGGCAAGTTCTT
CAACGGCTACATCGACACCGTCGCTGAAAGCATGAAGCTCTTCGACCACCTCGGTGGCCCCGA
5 GACCTACAACCACTACCCCATCGGGTGGGCAATGGCCTTCAACACCCCTACAAGCTGTTCAAG
CGCTACGCCTCGCATGAAGGCGGCATTGCCGACCCGGCAATCATCTCCTGGCCCAACGGCATT
GCCGCACACGGTGAAATCCGCGACAACCTACGTCAATGTCAGCGACATCACGCCCACCGTCTAC
GACCTGTTGGGCATGACACCGCCGGGGACCGTCAAGGGGATTCCGCAGAAACCGATGGACGG
CGTGAGCTTCATAGCGGCCCTTGCCGACCCGGCCGCCGACACCGGCAAGACCACCCAGTTCTA
10 CACCATGCTGGGCACCCGCGGGATCTGGCATGAAGGTTGGTTCGCCAACACCATTACGCGGC
CACGCCCCGCCGGCTGGTCAATTTCAACGCTGACCGCTGGGAACTGTTCCACATCGCAGCAGA
CCGCAGCCAGTGCCACGACCTGGCCGCCGAGCATCCCGACAACTTGAGGAGCTCAAGGCGCT
GTGGTTCTCCGAAGCCGCCAAGTACAACGGGCTGCCGCTGGCCGATCTGAACCTCCTGGAAAC
GATGACTCGGTGCGGGCCTTACCTGGTCAGCGAACGAGCCAGCTACGTCTACTATCCCGACTG
15 CGCTGACGTGGGCATCGGCGCGGCCGTAGAGATTCCGCGGGCGCTCGTTCGCCGTGCTGGCCG
ATGTGACCATCGATACCACCGGCGCCGAGGGCGTGCTGTTCAAGCACGGCGGCGCCCATGGC
GGGCACGTGCTGTTCTCGTCCGGGACGAGCGCTTGCACTACGTCTACAACCTCCTCGGTGAGCGC
CAGCAGCTGGTCAGCTCGTCGGGTCCGGTCCCGTCGGGAAGACATCTACTCGGGGTTCTGTTAT
TTGCGGACCGGAACCGTGCCCAACAGTCACACGCCGGTGGGCGATCTTGAGCTGTTCTTCGAC
20 GAGAACCTGGTCGGCGCCCTGACCAATGTGCTGACCCACCCTGGAACGTTCCGGTTGGCCGGC
GCCGCTATCAGCGTTGGCCGCAACGGCGGTTCCGGTGTGTCCAGCCACTACGAAGCGCCGTTT
GCGTTCACCGGCGGTACCATCACCCAGGTACCGTTCGACGTGTCAGGCCGACCGTTCAAGAT
GTGGAATCCGATCTTGCGCTTGCTTTTCGCGTGACTGA

25 >Rv0764c - lanosterol 14-demethylase cytochrome P450 TB.seq 856683:858035 MW:50879
>emb|AL123456|MTBH37RV:c858035-856680, Rv0764c SEQ ID NO:33
ATGAGCGCTGTTGCACTACCCCGGGTTTCGGGTGGCCACGACGAACACGGCCACCTCGAGGAG
TTCCGCACCGATCCGATCGGGCTGATGCAACGGGTCCGCGACGAATGCGGAGACGTCCGTACC
TTCCAGCTGGCCGGGAAGCAGGTCGTGCTGCTGTCCGGCTCGCACGCCAACGAATTCTTCTTC
30 CGGGCGGGCGACGACGACCTGGACCAGGCCAAGGCATACCCGTTTCATGACGCCGATCTTCGG
CGAGGGCGTGGTGTTCGACGCCAGCCCGGAACGGCGTAAAGAGATGCTGCACAATGCCGCGC
TACGCGGCGAGCAGATGAAGGGCCACGCTGCCACCATCGAAGATCAAGTCCGACGGATGATCG
CCGACTGGGGTGAGGCCGGCGAGATCGATCTGCTGGACTTCTTCGCCGAGCTGACCATCTACA
CCTCCTCGGCCTGCCTGATCGGCAAGAAGTTCCGCGACCAGCTCGACGGGCGATTCCGCCAAGC
35 TCTATCACGAGTTGGAGCGCGGCACCGACCCACTAGCCTACGTGACCCGTATCTGCCGATCG
AGAGCTTCCGTCGCCGCGACGAAGCCCGCAATGGTCTGGTGGCACTGGTTGCGGACATCATGA
ACGGCCGGATCGCCAACCCACCCACCGACAAGAGCGACCGTGACATGCTCGACGTGCTCATCG

CCGTCAAGGCTGAGACCGGCACTCCCCGGTTCTCGGCCGACGAGATCACCGGCATGTTTCATCT
CGATGATGTTCCGCCGCCATCACACCAGCTCGGGTACGGCTTCGTGGACGCTGATCGAGTTGA
TGCGCCATCGCGACGCCTACGCGGCCGTGATCGACGAACTCGACGAGCTGTACGGCGACGGC
CGATCGGTGAGTTTCCATGCGCTGCGCCAGATTCCGCAGCTGGAAAACGTGCTGAAAGAGACG
5 CTGCGCCTGCACCCTCCGCTGATCATCCTCATGCGAGTGGCCAAGGGCGAGTTCGAGGTGCAA
GGCCACCGGATTTCATGAGGGCGATCTGGTGGCGGCCCTCCCCGGCGATCTCCAACCGGATCCCC
GAAGACTTCCCCGATCCCCACGACTTCGTGCCAGCACGATACGAGCAGCCGCGCCAGGAAGAT
CTGCTCAACCGCTGGACGTGGATTCCGTTCCGGCGCCGGCCGGCATCGTTGCGTGGGGGCGGC
GTTCCGCATCATGCAGATCAAAGCGATCTTCTCGGTGTTGTTGCGCGAGTATGAGTTTGAGATG
10 GCGCAACCGCCAGAAAGCTATCGTAACGACCATTGCAAGATGGTGGTGCAGTTGGCCCAGCCC
GCTTGCGTGCGCTACCGCCGGCGAACGGGAGTTTAA

>Rv0861c - DNA helicase TB.seq 958524:960149 MW:59773

>emb|AL123456|MTBH37RV:c960149-958521, Rv0861c SEQ ID NO:34

15 GTGCAGTCCGATAAGACGGTGCTGTTGGAAGTCGACCATGAACTGGCCGGCGCTGCACGCGCC
GCCATCGCGCGGTTCCGCCGAGCTGGAACGTGCACCCGAACATGTCCACACCTACCGCATCACA
CCGCTGGCACTGTGGAATGCTCGCGCCGCCGGCCATGATGCCGAGCAAGTCGTGACGCGCT
GGTCAGTTACTCCCGCTACGCGGTGCCGCAACCCTTGCTCGTCGACATCGTCGACACCATGGC
CCGCTACGGACGACTGCAGTTGGTCAAGAACCCGGCCCATGGCCTGACGCTGGTGAGCCTGGA
20 CCGCGCGGTGCTTGAGGAAGTGCTGCGCAACAAGAAGATCGCGCCGATGCTTGCGCGCCCGCAT
CGATGACGACACCGTCGTGCTCCACCCAGCGAACGCGGCCGGGTCAAGCAGCTGCTGCTCAA
GATCGGTTGGCCCGCAGAGGATCTCGCCGGCTACGTCGATGGTGAAGCGCACCCGATCAGCCT
GCACCAGGAGGGCTGGCAGCTGCGCGATTACCAGCGGCTGGCCGCGGACTCGTTCTGGGCGG
GCGGCTCCGGGGTGGTGGTGCTGCCATGTGGGGCCGGCAAGACGCTGGTGGTGCGGGCCGC
25 AATGGCCAAAGCCGGCGCGACGACGTTGATCCTGGTCACCAATATCGTCGCGGCCCGGCAATG
GAAACGAGAGCTGGTCGCGCGCACCTCGCTCACCGAGAATGAGATCGGCGAATTCTCGGGAGA
ACGCAAGGAAATCCGACCTGTCACCATCTCGACATACCAGATGATCACCCGCCGCACTAAGGGC
GAGTACCGCCATCTGGAAGTGTTCGACAGCCGCGACTGGGGGCTCATCATCTATGACGAGGTG
CACCTGTTGCCGGCACCGGTCTTCCGGATGACCGCTGACCTGCAGTCCAAACGGCGGCTGGGG
30 CTGACCGCCACGTTGATCCGTGAAGACGGACGCGAGGGCGACGTGTTTTCCCTTATCGGACCA
AAGCGCTATGACGCGCCGTGGAAGGACATTGAGGCGCAGGGCTGGATCGCGCCAGCTGAGTG
CGTGGAAGTCCGGGTACGATGACCGACAGCGAGCGGATGATGTACGCCACCGCCGAACCCG
AAGAACGCTACCGGATCTGCTCGACGGTGACACCAAAATTGCTGTGGTCAAGTCGATTCTGGC
GAAGCACCCGGATGAGCAGACCCTGGTCATCGGAGCGTACTTGGATCAGCTCGACGAGCTGGG
35 CGCCGAGCTCGGCGCTCCGGTGATTAGGGGTGACAAAGGACCAGCGAACGCGAGGCACTGT
TCGACGCCTTCCGCCGCGGCGAGGTGCTACGCTCGTGGTGTCCAAGGTGGCTAACTTCTCCA
TCGACTTGCCGGAAGCCGCCGTGGCGGTACAGGTTTCGGGAACATTGGGCTCACGCCAGGAAG

AGGCGCAACGGCTCGGCCGGATATTGCGACCCAAGGCCGACGGGGCGGTGCCATCTTCTAC
TCGGTGGTGGCCCGGACAGCCTGGATGCCGAGTACGCCGCACACCGGCAGCGGTTTTAGCT
GAGCAGGGCTACGGTTACATCATCCGCGACGCCGACGACCTGCTGGGCCCGGCAATTTAG

5 >Rv0904c accD3 TB.seq 1006694:1008178 MW:51741

>emb|AL123456|MTBH37RV:c1008178-1006691, accD3 SEQ ID NO:35

GTGAGTCGTATCACGACCGACCAACTGCGGCACGCGGTGCTAGACCGGGGATCTTTCGTCAGC
TGGGATAGCGAGCCGCTGGCGGTGCCGGTAGCCGACTCCTATGCGCGGGAGCTGGCCGCCGC
TCGGGCGGCCACCGGCGCGGACGAATCGGTGCAGACCGGTGAGGGACGCGTATTCGGGCGG
10 CGGGTGGCCGTGGTGGCCTGTGAGTTCGACTTCCTGGGCGGCTCGATTGGGGTGGCAGCGGC
CGAACGGATCACCGCCGCCGTGAGCGGGCGACCGCCGAGCGGCTGCCGCTACTGGCGTCAC
CAAGCTCGGGAGGCACCCGCATGCAAGAAGGCACGGTCGCGTTTCTGCAGATGGTGAAGATCG
CTGCGGCCATCCAGCTGCACAACCAGGCGCGCCTGCCCTACCTGGTCTATTTGCGCCATCCGA
CCACGGGTGGAGTTTTCGCGTCTGGGGCTCGCTGGGGCATCTCACCGTCGCCGAGCCGGGC
15 GCCCTGATCGGCTTTCTGGGACCACGGGTCTATGAGTTGCTCTATGGCGACCCCTTCCCATCCG
GCGTCCAAACCGCCGAGAATCTACGGCGGCATGGGATCATCGACGGCGTCGTTGCACTGGACC
GGCTACGACCGATGCTGGATCGTGCCTTGACGGTGCTCATCGACGCTCCCGAACCGCTTCCGG
CACC GCAGACGCCCGCGCCCGTACCCGATGTGCCACGTGGGACTCGGTGGTGGCATCGCGC
CGGCCCGACCGGCCGGCGTCAGGCAGCTACTGCGACACGGCGCCACCGACCGGGTGTGTT
20 GTCAGGAACCGATCAAGGCGAAGCGGCGACCACGCTGCTGGCGCTGGCCCGCTTTGGCGGCC
AACCCACGGTGGTCCTCGGCCAGCAAAGGGCAGTAGGCGGCGGGGAAGCACTGTCGGGGCC
GCTGCGTTACGCGAAGCCCGACGCGGGATGGCGCTCGCCGCCGAGCTGTGCCTGCCGCTGGT
GCTGGTCATTGACGCGGCCGGACCCGCGTTGTGCGCCGCAGCCGAACAGGGCGGGCTGGCCG
GCCAGATCGCGCATTGCCTGGCCGAGCTCGTCACGCTGGATACCCGACCGTGTGATCCTGC
25 TGGGCCAGGGCAGCGGCGGGCCGGCGTGGCGATGTTGCCCGCCGACCGGGTGGTGGCCGC
ACTCCACGGCTGGCTGGCGCCCTTGCCCTCCCGAAGGAGCCAGCGCGATCGTGTTCCGAGACAC
TGCTCATGCCGCCGAACCTCGCTGCCGCCCAAGGCATCCGGTCCGCCGACCTACTGAAGTCGGG
GATTGTCGACACCATCGTGCCGGAGTACCCCGACGCCGCAGACGAGCCGATCGAGTTCGCCCT
ACGACTGTGGAACGCCATCGCCGCCGAAGTGACGCGTTACGGAAGATACCGGCCCCCGGAACG
30 CCTCGCGACTCGGTTGCAACGCTACCGCCGGATCGGGTTGCCCCGCGACTAA

>Rv0983 - TB.seq 1099064:1100455 MW:46454

>emb|AL123456|MTBH37RV:1099064-1100458, Rv0983 SEQ ID NO:36

ATGGCCAAGTTGGCCCCGAGTAGTGGGCCTAGTACAGGAAGAGCAACCTAGCGACATGACGAAT
35 CACCCACGGTATTCGCCACCGCCGCAGCAGCCGGGAACCCAGGTTATGCTCAGGGGCAGCA
GCAAACGTACAGCCAGCAGTTGCACTGGCGTTACCCACCGTCCCCGCCCCCGCAGCCAACCCA
GTACCGTCAACCCTACGAGGCGTTGGGTGGTACCCGGCCGGTCTGATACCTGGCGTGATTCC

GACCATGACGCCCCCTCCTGGGATGGTTGCGCAACGCCCTCGTGACGGCATGTTGGCCATCGG
CGCGGTGACGATAGCGGTGGTGTCCGCCGGCATCGGCGGCGCGGCCGCATCCCTGGTCGGGT
TCAACCGGGCACCCGCCGGCCCCAGCGGCGGCCAGTGGCTGCCAGCGCGGCGCCAAGCAT
CCCCGCAGCAACATGCCGCCGGGTTCGGTTCGAACAGGTGGCGGCCAAGGTGGTGCCCAAGTG
5 TCGTCATGTTGGAACCGATCTGGGCCGCCAGTCGGAGGAGGGCTCCGGCATCATTCTGTCTG
CCGAGGGGGCTGATCTTGACCAACAACCACGTGATCGCGGCGGCCGCCAAGCCTCCCCTGGGC
AGTCCGCCGCCGAAAACGACGGTAACCTTCTCTGACGGGCGGACCGCACCCCTTCACGGTGGTG
GGGGCTGACCCACCAGTGATATCGCCGTCGTCCGTGTTACAGGGCGTCTCCGGGCTCACCOCG
ATCTCCCTGGGTTCTCCTCGGACCTGAGGGTCGGTCAGCCGGTGCTGGCGATCGGGTCGCCG
10 CTCGGTTTGAGGGGCACCGTGACCACGGGGATCGTCAGCGCTCTCAACCGTCCAGTGTCGACG
ACCGGCGAGGCCCGCAACCAGAACACCGTGCTGGACGCCATTAGACCGACGCCGCGATCAA
CCCCGGTAACCTCCGGGGGCGCGCTGGTGAACATGAACGCTCAACTCGTCGGAGTCAACTCGGC
CATTGCCACGCTGGGCGCGGACTCAGCCGATGCGCAGAGCGGCTCGATCGGTCTCGGTTTTGC
GATTCCAGTCGACCAGGCCAAGCGCATCGCCGACGAGTTGATCAGCACCGGCAAGGCGTCACA
15 TGCCTCCCTGGGTGTGCAGGTGACCAATGACAAAGACACCCTGGGCGCCAAGATCGTCGAAGT
AGTGGCCGGTGGTGTGCGCGAACGCTGGAGTGCCGAAGGGCGTCGTTGTACCAAGGTGCG
ACGACCGCCCGATCAACAGCGCGGACGCGTTGGTTGCCGCCGTGCGGTCCAAAGCGCCGGGC
GCCACGGTGCGGCTAACCTTTCAGGATCCCTCGGGCGGTAGCCGCACAGTGCAAGTCACCCTC
GGCAAGGCGGAGCAGTGA

20

>Rv1008 - Similar to E.coli protein YcfH TB.seq 1127087:1127878 MW:29066

>emb|AL123456|MTBH37RV:1127087-1127881, Rv1008 SEQ ID NO:37

TTGGTGCACGCCCCACCCATCTCGACGCGTGCGGTGCACGAGACGCCGATACGGTGCGGTC
GCTCGTCGAGCGAGCCGCCGCGGCCGGCGTGACCGCGGTGGTCACCGTCGCCGACGACCTG
25 GAGTCCGCGCGCTGGGTACCCGCGCGGCCGAATGGGATCGGCGAGTCTATGCCGCGGTGGC
GTTGCACCCGACCCGCGCCGATGCGCTCACCAGCGCTGCCCGTGCCGAGCTCGAGCGATTGG
TTGCCACCCACAGGGTGGTGGCCGTGGTGAGACCGGAATCGACATGTACTGGCCGGGTGCG
CTGGACGGGTGTGCGGAGCCGCACGTCCAGCGGGAGGCCTTTGCCTGGCATATCGATCTGGC
CAAGCGGACCGGTAACCGCTGATGATCCACAATCGTCAGGCCGACCGCGACGTGCTGGACGT
30 GCTGCGGGCCGAGGGCGCGCCGGACACCGTGATCTTGCACTGCTTCTCGTCGGACGCGGCGA
TGGCCCGCACGTGTGTGGACGCCGGGTGGCTGCTCAGCCTGTCCGGGACGGTGAGCTTCCGT
ACCGCCCGTGAACTACGGGAAGCCGTCCCGCTGATGCCGGTGGAGCAGCTTTTGGTGGAAACC
GATGCACCGTATTTGACCCCGCATCCCCACCGGGGCTTGCGGAACGAACCGTACTGCCTGCC
TATACCGTGCGGGCGCTGGCTGAACTGGTCAATCGGCGCCCCGAAGAGGTGGCGCTCATCACC
35 ACAAGCAACGCTCGCCGAGCTTATGGGCTAGGGTGGATGCGCCAATGA

>Rv1009 - lipoprotein, similar to various other MTB proteins TB.seq 1128089:1129174 MW:38079

>emb|AL123456|MTBH37RV:1128089-1129177, Rv1009 SEQ ID NO:38

ATGTTGCGCCTGGTAGTCGGTGGCTGCTGCTGGTGTGGCGTTCGCCGGTGGCTATGCGGTC
GCCGCATGCAAAACGGTGACGTTGACCGTCGACGGAACCGCGATGCGGGTGACCACGATGAAA
5 TCGCGGGTGATCGACATCGTCGAAGAGAACGGGTTCTCAGTCGACGACCGCGACGACCTGTAT
CCCGCGGCCGGCGTGCAGGTCCATGACGCCGACACCATCGTGCTGCGGCGTAGCCGTCCGCT
GCAGATCTCGCTGGATGGTCACGACGCTAAGCAGGTGTGGACGACCGCGTCGACGGTGGACG
AGGCGCTGGCCCAACTCGCGATGACCGACACGGCGCCGGCCGCGGCTTCTCGCGCCAGCCGC
GTCCCGCTGTCCGGGATGGCGCTACCGGTCGTACGCCCAAGACGGTGCAGCTCAACGACGG
10 CGGGTTGGTGCACGCGGTGCACTTGCCGGCCCCCAATGTGCGGGGGCTGCTGAGTGCGGGCCG
GCGTGCCGCTGTTGCAAAGCGACCACGTGGTGCCCGCCGCGACGGCCCCGATCGTCGAAGGC
ATGCAGATCCAGGTGACCCGCAATCGGATCAAGAAGGTCACCGAGCGGCTGCCGCTGCCGCCG
AACCGCGCTCGTGTCGAGGACCCGGAGATGAACATGAGCCGGGAGGTGCTCGAAGACCCGGG
GGTTCGGGGGACCCAGGATGTGACGTTGCGGGTAGCTGAGGTCAACGGCGTCGAGACCCGGCC
15 GTTTGCCCGTCGCCAACGTCGTGGTGACCCCGGCCACGAAGCCGTGGTGCGGGTGGGCACC
AAGCCCGGTACCGAGGTGCCCCCGGTGATCGACGGAAGCATCTGGGACGCGATCGCCGGCTG
TGAGGCCGGTGGCAACTGGGCGATCAACACCGGCAACGGGTATTACGGTGGTGTGAGTTTGA
CCAGGGCACCTGGGAGGCCAACGGCGGGCTGCGGTATGCACCCCGCGCTGACCTCGCCACCC
GCGAAGAGCAGATCGCCGTTGCCGAGGTGACCCGACTGCGTCAAGGTTGGGGCGCCTGGCCG
20 GTATGTGCTGCACGAGCGGGTGGCGCTGA

>Rv1010 ksgA 16S rRNA dimethyltransferase TB.seq 1129150:1130100 MW:34647

>emb|AL123456|MTBH37RV:1129150-1130103, ksgA SEQ ID NO:39

ATGTGCTGCACGAGCGGGTGCGCGCTGACCATCCGGCTGCTCGGGCGCACTGAGATCAGGCC
25 GCTGGCCAAAGAGCTCGACTTTCGGCCGCGCAAATCTCTCGGACAGAACTTCGTGCACGACGC
CAACACGGTGCGACGGGTGGTTGCCGCCTCCGGGGTCAGCCGTTCCGACCTGGTTTTGGAGGT
CGGGCCGGGCTGGGATCGCTGACCCTGGCACTGCTCGACCGCGGCGGACCGTCACCGCGG
TCGAGATCGATCCACTACTGGCTTCTCGGCTGCAACAGACCGTGGCGGAGCACTCGCACAGCG
AGGTTACCGACTAACGGTGGTCAATCGCGACGTCTGGCCCTGCGCCGGGAGGATCTAGCCG
30 CGGCGCCGACCGCGGTGGTTGCCAATCTGCCGTACAACGTAGCGGTACCGGCGTTGTTGCATC
TGCTTGTGAGTTCCCGTCGATCCGTGTCGTGACGGTGATGGTGCAGGCCGAGGTCGCCGAAC
GGCTCGCCGCCGAGCCGGGCAGCAAAGAGTACGGCGTGCCGAGCGTTAAGCTGCGCTTCTTC
GGGCGGGTTCCCGCTGCGGCATGGTGTGCGCCGACCGTTTTCTGGCCCATTCGCGGTGTCTAT
TCCGGGCTGGTACGCATCGATCGATATGAGACCTCGCCCTGGCCACCGACGACGCTTTTCGA
35 CGGCGGGTATTGAACTCGTGGACATCGCATTGCGCGAGCGGCGCAAGACTTCTCGCAACGCG
TTTGTGAGTGGGCGGGCTCGGAAGCGAGTCGGCGAATCGATTGTTGGCGGCCAGCATCGAC
CCCGCCCGTCGCGGTGAGACGCTGTCCATCGACGACTTCGTGCGGCTGCTGCGACGGTCCGG

CGGCTCCGACGAGGCCACCAGCACCGGCCGGGACGCCAGGGCGCCGGACATTTGCGGGGCAC
GCGTCGGCGAGCTGA

>Rv1011- Homology to E.coli protein YcbH TB.seq 1130189:1131106 MW:31350

5 >emb|AL123456|MTBH37RV:1130189-1131109, Rv1011 SEQ ID NO:40

GTGCCCACCGGGTCGGTCACCGTTCCGGGTGCCCGGAAAGGTCAACCTCTATCTGGCGGTCCGC
GATCGCCGCGAGGACGGCTATCACGAGCTGACCACGGTATTTTCATGCCGTCTCGCTGGTCGAC
GAGGTAACCGTTCGTAACGCTGATGTGCTCTCGCTCGAGTTGGTCGGCGAGGGGGCCGACCAG
CTGCCGACCGACGAACGCAATCTCGCCTGGCAGGCGGCCGAGCTGATGGCCGAACACGTGGG
10 CCGGGCGCCGGACGTCTCGATCATGATCGACAAATCCATTCCGGTCGCCGGCGGCATGGCCG
GTGGCAGCGCGGACGCTGCGGCGGTCTGTTGCGATGAACTCGTTGTGGGAACTCAATGTGC
CCCGCCGCGACCTGCGCATGCTCGCCGCGCGGCTAGGCAGCGATGTGCCGTTTGCCCTGCAT
GGTGGTACCGCGCTGGGGACGGGTGCGGCGGAGGAGTTGGCCACCGTGTTATCCCGCAACAC
CTTCCACTGGGTCTGCGCTTCGCCGACAGCGGGTTGCTCACCTCCGCGGTGTACAACGAGCT
15 CGACCGGCTCAGGGAGGTGGGGGATCCGCCCGGCTTGGTGAGCCCGGGCCGGTTCTGGCTG
CCTTAGCTGCGGGTGATCCGGATCAGCTGGCGCCGTTGCTGGGTAATGAAATGCAAGCGGCCG
CGGTGAGCCTGGACCCGGCGCTGGCTCGTGCGTTACGCGCCGGTGTGGAGGCCGGCGCGCTC
GCAGGCATCGTGCCGGTTCGGGTCCCACGTGTGCCTTCCTGTGCACCTCGGCGAGCTCGGCG
ATCGATGTCGGCGCGCAGCTGTCGGGGGCGGGAGTTTGTGCGACCGTTCGAGTCGCCACCGG
20 GCCGGTACCGGCGCCCGCGTGGTGTCTGCGCCGACCGAAGTGTGA

>Rv1106c - cholesterol dehydrogenase TB.seq 1232845:1233954 MW:40743

>emb|AL123456|MTBH37RV:c1233954-1232842, Rv1106c SEQ ID NO:41

ATGCTTCGCCGCATGGGTGATGCATCGCTGACAACCGAGCTCGGCCGCGTTCTGGTCACCGGC
25 GGCGCGGGCTTCGTGGGCGCCAACCTGGTGACCACCTTGCTGGACCGCGGGCACTGGGTGCG
TTCCTTCGACCGCGCGCCGTCGCTGTTGCCTGCGCATCCGCAACTGGAGGTGCTGCAAGGGGA
CATCACCGACGCGGACGTCTGCGCCGCGGCCGTGGACGGCATCGACACGATCTTCCACACCG
CAGCGATCATCGAGCTGATGGGCGGCGCGTCCGTACCGACGAGTACCGCCAACGTAGCTTTG
CGGTCAACGTGCGCGGCACCGAGAACCTGCTGCACGCGGCCAGCGGGCCGGGTGCAGCG
30 GTTCGTCTACACGTATCCAACAGTGTGGTGATGGGCGGCCAGAACATCGCCGGCGGTGACGA
GACGCTGCCCTATACCGACCGGTTCAACGACCTCTACACCGAGACCAAGGTGGTTGCCGAGCG
ATTCGTGTTGGCCCAGAACGGTGTGACGGCATGCTGACGTGCGCGATCCGGCCCAGCGGCAT
CTGGGGAAACGGCGATCAGACGATGTTCCGCAAGCTGTTGAAAAGTGTGCTCAAGGGCCACGT
CAAGGTGCTGGTGGGCGCAAGTCGGCCCGGCTGGATAACTCTTACGTGCACAACCTGATTCA
35 CGGTTTCATCTTGGCCGCTGCCCATCTGGTGCCGGACGGCACAGCGCCCGGGCAGGCTTACTT
CATCAACGACGCAGAGCCGATCAATATGTTGAGTTCGCTCGGCCGGTGCTCGAGGCGTGCGG
GCAGCGCTGGCCGAAGATGCGGATTTCCGGCCCCGCGGTCCGCTGGGTAATGACGGGGTGGC

AGCGGCTGCACTTCCGGTTCGGATTCCCCGCGCCGCTGCTCGAGCCGCTGGCCGTGGAACGAC
TGTACCTGGACAACACTACTTTTCGATCGCTAAGGCACGCCGCGACCTGGGCTATGAGCCGCTGTT
CACCACCCAGCAGGCGCTGACCGAATGCCTGCCGTACTACGTGAGTCTGTTTGAGCAGATGAA
GAACGAGGCCCGGGCGGAAAAAACGGCCGCCACAGTCAAGCCGTAG

5

>Rv1110 lylB2 TB.seq 1236183:1237187 MW:36298

>emb|AL123456|MTBH37RV:1236183-1237190, lylB' SEQ ID NO:42

ATGGTTCCGACGGTCGACATGGGGATTCCCGGGGCTTCGGTATCGTCGCGATCGGTGGCCGAG
CGTCCCAACCGTAAGCGGGTGCTGCTGGCCGAGCCGCGTGGCTACTGCGCTGGCGTGGATCG
10 GGCCGTGGAACGGTGAACGCGCGCTTCAAAAACACGGCCCGCCTGTCTACGTGCGTCACGA
GATCGTGATAACCGCCACGTGGTTGACACCTGGCTAAGGCCGGTGCGGTTTTCTGTCGAAGA
GACCGAGCAGGTTCCCGAGGGAGCGATTGTGGTGTCTCCGCGCACGGGGTCGCGCCTACGG
TGCACGTCAGCGCCAGCGAGCGCAACCTGCAGGTCATTGACGCCACCTGCCCCGCTGGTCACCA
AGGTGCACAACGAGGCCAGGCGGTTCCGCCGGGACGACTACGACATCTTGCTGATCGGTCATG
15 AGGGCCACGAGGAAGTCGTCGGTACTGCTGGGGAAGCTCCCGATCATGTGCAGCTGGTCGACG
GGGTGGACGCCGTCGACCAGGTGACCGTCCGTGACGAGGACAAAGTGTTTTGGCTGTCGCAG
ACCACCTGTCCGTCGATGAGACCATGGAGATTGTCGGGCGGTTGCGTCGGCGTTTCCCCAAG
CTGCAGGATCCGCCAGCGACGACATCTGCTATGCGACCCAGAATCGGCAGGTGCGGGTCAAG
GCGATGGCGCCCCGAGTGCGAGCTGGTCATCGTGGTCCGGCTCGCGCAATTGTCGAATTCGGTT
20 CGGCTGGTCGAGGTGGCGCTGGGTGCCGGGGCGCGGGCCGCCACCTGGTGGACTGGGCGG
ACGATATCGACTCGGCCTGGCTGGACGGCGTTACCACGGTCGGCGTTACGTCGGGGGCATCGG
TCCCCGAGGTGCTGGTGCGCGGTGTGCTGGAGCGGCTGGCCGAATGCGGCTACGACATCGTG
CAACCGGTGACAACGGCCAACGAGACGTTGGTGTTCGCATTGCCCGGGAGCTCCGCTCACCT
CGCTGA

25

>Rv1216c - TB.seq 1359473:1360144 MW:24863

>emb|AL123456|MTBH37RV:c1360144-1359470, Rv1216c SEQ ID NO:43

ATGCACATTGGGCTGAAGATATTCATATGGGGCGTGTTAGGACTCGTCGTTTTCGGCGCGCTCC
TATTCGGGCCAGCCGGCACGTTCCGACTATTGGCAGGCGTGGGTGTTCTCGCCGCATTTGTGA
30 GCACCACGATTGGCCCCACAATCTATCTGGCTCGCAACGATCCCGCGGCCCTTCAACGTGCGAT
GCGCAGCGGTCCGCTCGCGGAGGGCCGAACGATTGAGAAGTTCATCGTCATCGGCGCTTTTCT
GGGGTTCTTCGCGATGATGGTGCTGAGCGCGTGCGACCATCGTTATGGTTGGTTCGTGAGTCC
AGCCGCGGTGTGCGTGATCGGCGACGTCTAGTGATGACGGGCCCTTGGCATCGCCATGCTGGT
GGTCATCCAGAACAGGTATGCCGCCTCGACGGTCAGGGTGAGGGCGGGCCAGATATTGGCCTC
35 CGACGGTCTCTACAAATTGTCCGACACCCGATGTACGCCGGGAACGTGGTCATGATGACAGG
CATACCGCTGGCACTGGGCTCTTACTGGGCGATGTTATCCTCGTCCCCGGCACACTGGTGTTG

GTGTTCCGCATCCTCGACGAGGAAAACTACTGACGCAAGAACTCAGCGGGTACCGCGAATACC
GGCAACTGGTGCGCTACCGGTTGGTGCCCTACGTGTGGTAG

>Rv1223 htrA TB.seq 1365810:1367456 MW:56547

5 >emb|AL123456|MTBH37RV:1365810-1367459, htrA SEQ ID NO:44

GTGAGCCACTTGTGCGCAGCGCATGGCGGGGTTGCTGCCAGTTCATGGCGAGTGGTCGCGATCC
GTGGATACTAGGGTGGACACGGACAACGCGATGCCTGCACGTTTTAGCGCCCAGATTCAGAAT
GAGGATGAGGTGACCTCCGACCAAGGCAACAACGGCGGCCCGAACGGCGGAGGCCGCCTGGC
GCCGCGCCCCGGTTTTTCGGCCACCGGTGACCCGGCGTCGCGTCAAGCGTTCGGGCGTCCGT
10 CCGGGGTCCAAGGGTCCCTTTGTGGCCGAGCGTGTGCGCCCGCAGAAGTACCAGGACCAGTCT
GACTTCACACCGAACGATCAGCTTGCTGACCCGGTGCTTCAGGAGGCGTTCGGTTCGTCCGTTT
GCGGGCGCCGAATCGCTGCAGCGCCATCCCATCGATGCCGGAGCGCTGGCAGCTGAGAAAGA
CGGTGCCGGCCCCGACGAGCCCGACGATCCGTGGCGCGACCCCGCGGCCGCGGCCGCGCTG
GGGACGCCAGCGCTAGCCGCGCCGGCACCGCACGGTGCGCTGGCCGGCAGCGGCAAGCTGG
15 GTGTGCGCGACGTGCTGTTTGGCGGCAAGGTGTCCTACTTGGCGCTGGGCATCTTGGTCGCTA
TCGCACTGGTGATCGGCGGCATCGGCGGTGTGTCGCGCCGAAGACCGCGGAAGTAGTCGAT
GCGTTCACCACGTGCAAGGTGACCCTGTGACCACTGGCAATGCCAGGAACCGGCCGCGCG
GTTCCACCAAGGTGGCGGCCCGCGTGGCCGATTGCGTGGTGACCATGAGTCGGTCAGCGACCA
GGAGGGCATGCAAGGTTCCGGCGTCATCGTCGATGGCCGCGGCTACATCGTCACCAACAATCA
20 CGTGATCTCTGAGGCGGCCAACAAATCCCAGCCAGTTCAGACGACCGTGGTGTCAACGACGG
CAAGGAGGTGCCCGCCAATCTGGTGGGTGCTGACCCCAAGACCGACTTGGCCGTCCTCAAGGT
CGACAACGTGACAATCTGACCGTGGCCCGGCTCGGTGATTCCAGCAAGGTACGGGTGCGTGA
CGAAGTCCTCGCGGTGCGCGCGCCCTGGGGCTGCGCAGTACGGTGACCCAGGGCATTGTCA
GCGCGCTACACCGCCCCGTTCCGTTGTGCGGCGAGGGCTCTGACACCGACACCGTCAATTGACG
25 CAATTCAGACCGACGCCTCGATCAACCACGGTAACTCCGGCGGTCCGCTAATCGACATGGATGC
CCAGGTGATTGGCATCAACACCGCCGGTAAGTCACTGTGCGATAGCGCCAGCGGGCTGGGCTT
TGCGATCCCGGTCAACGAGATGAAATTGGTGGCAAATCTCTGATCAAAGACGGAAAGATCGTG
CATCCGACGTTGGGCATCAGCACCCGGTCAGTAAGCAACGCGATCGCGTCGGGCGCGCAGGT
GGCCAATGTAAAGGCGGGAAGTCCCGCGCAGAAGGGCGGGATCTTGGAGAACGATGTGATCGT
30 CAAGGTGCGTAACCGCGCGGTGCGCGACTCCGACGAGTTCGTGCTGCGCGTCCGTCGCCAGTTGG
CTATCGGCCAGGACGCTCCGATAGAGGTGGTCCGCGAGGGTCGGCATGTGACGCTGACGGTG
AAACCGGACCCCGATAGCACCTAG

>Rv1224 - TB.seq 1367461:1367853 MW:14083

35 >emb|AL123456|MTBH37RV:1367461-1367856, Rv1224 SEQ ID NO:45

GTGTTCCGCCAACATCGGTTGGTGGGAAATGCTCGTCCTCGTCATGGTCCGGCTGGTGGTGCTT
GGCCCGGAGCGGCTCCCGGGTGCCATCCGCTGGGCGGCAAGCGCTCTGCGGCAGGCGCGCG

ACTATCTCAGCGGTGTGACCAGCCAGCTACGTGAGGACATTGGACCCGAATTCGATGATCTGCG
GGGACATCTCGGTGAGCTGCAGAAAGCTACGGGGAATGACTCCGCGGGCTGCGTTGACCAAGCA
CCTACTGGATGGCGATGATTCCCTGTTACCCGGAGACTTCGACCGACCGACGCCGAAGAAACC
GGATGCGGGCGGGCTCGGCGGGGCCGGACGCTACTGAGCAGATCGGTGCGGGGGCCCATCCCG
5 TTTGACAGCGATGCCACCTAG

>Rv1229c mrp similar to MRP/NBP35 ATP-binding proteins TB.seq 1371778:1372947 MW:41064

>emb|AL123456|MTBH37RV:c1372947-1371775, mrp SEQ ID NO:46

ATGCCAAGCCGCCTACACTCGGCGGTGATGTCCGGAACCTCGTGATGGCGACCTGAACGCGGCG
10 ATACGCACCGCGCTGGGCAAGGTAATCGACCCCGAATTGCGGCGCCCCATCACCGAACCTGGGG
ATGGTCAAAAGCATCGACACCGGCCCGGATGGGAGCGTGACGTCGAGATCTACCTGACCATC
GCCGGCTGCCCGAAGAAGTCCGAAATCACCGAGCGTGTCACCCGGGCGGTGCGCCGACGTGCC
AGGCACTTCGGCGGTGCGGGTCAGCTTGGACGTGATGAGCGACGAGCAGCGCACCGAGCTGC
GTAAGCAGTTGCGTGCGGATACCCGCGAACCCGTATCCCGTTGCGCGAACCCGATTCTTTGAC
15 CCGGGTGTATGCCGTGGCTTCCGTAAGGGCGGAGTCGGAAAGTCCACCGTCACGGTCAACCT
GGCCGCCGCGATGGCCGTCCGCGGCCCTGTCGATCGGGGTGCTGGACGCTGATATCCACGGCC
ACTCTATCCCCCGATGATGGGCACACCGACCGGCCCTACCCAGGTTGAGTCGATGATCCTGC
CGCCGATCGCCACCAGGTGAAGGTCATCTCGATAGCCAGTTACCCAGGGCAACACCCCGG
TGGTGTGGCGCGGGCCGATGCTGCACCGGGCGTTGCAGCAGTTTCTGGCCGACGTGTAAGTGG
20 GGGGATCTGGACGTGCTGCTGCTGGACTTGCCGCCCGGAACCGGCGACGTCGCCATCTCGGT
GGCTCAACTGATCCCCAACGCCGAACCTCCTGGTGGTCAACACCCCGCAGCTGGCCGCCGCGGA
GGTGGCCGAACGGGCCGGCAGCATCGCGCTGCAAACCCGCCAACGCATCGTCGGCGTCGTGG
AGAACATGTCGGGGCTCACGCTGCCGGACGGCACACGATGCAGGTGTTGCGCGAGGGCGGT
GGCCGGCTGGTCGCCGAGCGGTTGTCGCGTGCGGTGCGCGCCGACGTGCCGCTGCTGGGTCA
25 GATCCCGCTGGACCCCGCACTGGTGGCCGCCGGCGATTGCGGCGTACCGCTCGTGTTGAGCT
CGCCGGAATCGGCGATCGGCAAGGAAGTGCATAGCATCGCCGACGGCTTGTGCACTCGACGAC
GCGGATTGGCGGGCATGTCGCTGGGGTTGGACCCGACACGACGCTAG

>Rv1239c corA magnesium and cobalt transport protein TB.seq 1381943:1383040 MW:41470

30 >emb|AL123456|MTBH37RV:c1383040-1381940, corA SEQ ID NO:47

GTGTTCCAGGGTTTGACGCATTGCCCGAAGTGCTGCGACCGGTGCGCGACCCAGCCGCCG
AACGCACACCCCGTTGCCAGCCACCGGCCCAAGCCTTGGTCGACTGCGGTGTCTACGTCTGC
GGCCAGCGACTGCCCGGCAAGTACACCTACGCCGCCGCGCTGCGCGAGGTGCGCGAGATCGA
ACTGACCGGGCAGGAGGCGTTCTGCTGGATCGGGCTGCACGAGCCCGATGAAAACAGATGCA
35 GGACGTAGCAGACGTTTTCGGGTTGCACCCGTTAGCCGTTGAGGACGCCGTGCACGCGCACCA
GCGACCCAAGTTGGAGCGCTACGACGAGACGCTGTTCTCTGTCCTCAAGACCGTCAACTACGT
CCCGCACGAATCGGTGGTACTGGCCCGCGAGATCGTCAAAACCGGCGAGATCATGATCTTCGT

CGGCAAGGATTTCTGGTACCGTCCGCCACGGCGAACACGGCGGGTTATCCGAGGTGCGTAA
GCGGATGGATGCCGACCCCGAACATTTGCGGTTGGGACCGTATGCGGTGATGCACGCGATCGC
CGACTACGTGGTTCGACCACTACCTCGAGGTGACCAATCTCATGGAGACCGATATCGACAGCATC
GAGGAAGTAGCGTTTCGCGCCGGGCGCAAGCTCGACATCGAACCGATCTATCTGCTCAAGCGG
5 GAAGTGGTTCGAGTTGCGCCGGTGCCTGAATCCGCTATCGACCGCATTCCAGCGCATGCAGACC
GAGAGCAAAGACCTCATTTGAAAGAAGTGGCGGCTACCTGCGCGACGTGCGCGACCAACCAG
ACCGAGGCCCGCGACCATGCGCCAGCTACGACGACATGCTCAACTCGCTGGTGCAGGCCGC
GCTCGCCCGGGTGGCATGCAGCAAAACATGGACATGCGCAAGATATCCGCGTGGGCAGGTAT
CATCGCGGTCCCCACCATGATCGCGGGCATCTATGGCATGAACTTTCACTTCATGCCCGAGCTG
10 GACTCCAGGTGGGGTTACCCGACAGTGATCGGCGGGATGGTCCTTATCTGTCTGTTCTCTACC
ACGTCTTCCGCAACAGAACTGGCTCTAG

>Rv1279 - TB.seq 1430060:1431643 MW:57332

>emb|AL123456|MTBH37RV:1430060-1431646, Rv1279 SEQ ID NO:48

15 ATGGACACTCAGAGCGACTACGTCTGGTTCGGTACCGGCTCAGCCGGGGCGGTTGTGGCCAG
CCGGCTTAGCACCGATCCGGCCACGACGGTGGTGGCCCTGGAGGCGGGGCGCGTGACAAGA
ACAGATTCATCGGCGTCCCAGCGGCGTTTTCCAAGCTGTTCCGCAGCGAGATCGACTGGGATTA
CCTAACCGAACCGCAGCCGGAGCTCGACGGCGCGAAATCTATTGCGCTCGTGGAAGGTGCT
CGGTGGCTCGTCTCATGAACGCAATGATGTGGGTGCGTGGATTTCGCATCAGACTACGATGA
20 GTGGGCCGCGCGAGCCGGTCCGCGGTGGTCTGACGCCGACGTGCTCGGCTACTTTGCGCGCA
TCGAGAACGTCACCGCTGCCTGGCACTTTGTGACGGTGACGACAGCGGAGTAACCGGTCCGT
TGCATATTTCCCGGCAACGCAGCCCAAGATCGGTGACCGCAGCGTGGCTGGCAGCCGCACGTG
AGTGCGGATTTGCCGCTGCGCGGCCGAATCCCTCGACCGGAAGGCTTTTGCGAGACCGTCG
TCACCCAGCGCCGCGGTGCTCGATTCACTACTGCCGACGCTATCTGAAGCCCGCGATGCGCC
25 GTAAAAACCTCCGTGTGCTTACCGGCGCCACTGCTACCCGGGTGGTCATCGACGGCGACCGGG
CCGTCCGGCGTGAATACCAAAGCGACGGTCAAACCCGCATCGTCTACGCCCGCCGCGAGGTG
GTGCTCTGCGCTGGTGGCTCAACAGCCCTCAGCTGCTGATGCTCTCCGGCATCGGCGACCGC
GACCACCTCGCCGAACACGACATCGACACCGTTTACCACGCGCCCGAGGTGGGTGCAACCTG
CTCGATCATCTCGTCACGGTGTGGGTTTCGACGTGAAAAGGACAGCTTGTTTGCCGCCGAGA
30 AGCCCGGCCAGTTGATCAGCTACTTACTGCGACGCCGCGGCATGCTCACCTCCAACGTGCGCG
AGGCGTACGGATTTGTCCGACGCCGACCCGAAGTGAAGCTGCCCCGATTGGAGTTGATTTTGC
CCCGGCCCGGTTTTACGACGAAGCGCTGGTTCCACCGGCTGGTCACGGTGTGGTATTCGGCCCC
GATTCTGGTCGCGCCGCAAGCCGTGGCCAGATCACGCTGCGGTCCGCCGATCCGCATGCCAA
GCCTGTATCGAACCGCGTTACCTGTCCGATCTCGGTGGCGTAGACCGGGCCGCCATGATGGC
35 GGGCCTGCGGATATGCGCGCGGATCGCGCAGGCCCGCCCGCTCAGAGATCTCCTTGGGTCCA
TCGCGCGACCGCGCAACAGCACCAGCTGGACGAGGCCACTCTCGAGTTGGCGCTGGCCACT
TGTTGCGACACCCTGTACCACCCGATGGGCACCTGCCGCATGGGCAGCGACGAGGCCAGCGT

GGTGGATCCGCAGCTGCGGGTCCGCGGTGTCGACGGACTCCGCGTCGCCGACGCGTCGGTGA
TGCCCAGCACGGTTCGTGGGCATACGCATGCGCCGTCGGTGCTGATCGGGGAGAAGGCCGCC
GACTTAATCCGCAGCTGA

- 5 >Rv1294 thrA homoserine dehydrogenase TB.seq 1449373:1450695 MW:45522
>emb|AL123456|MTBH37RV:1449373-1450698, thrA SEQ ID NO:49
GTGCCCCGGTGACGAAAAGCCGGTCCGGCTAGCGGTACTCGGTTTGGGCAACGTCGGCAGCGA
GGTTGTCCGCATCATCGAGAACAGCGCCGAGGATCTCGCGGCTCGTGCGGTGCCCCATTGGT
CCTGCGGGGCATCGGCGTGCGCCGCGTGACGACCGATCGCGGCGTGCCGATCGAATTGTTGA
10 CCGACGACATTGAAGAGCTCGTGGCCCCGCGAGGATGTCGATATCGTGGTGGAAGTGATGGGGC
CGGTGGAACCGTCGCGCAAGGCGATCCTGGGCGCCCTTGAGCGCGGCAAGTCCGTGCTTACG
GCGAACAAGGCTTTACTCGCCACCTCCACCGGCGAATTGGCACAGGCCGCCGAAAGCGCCCAT
GTTGATCTGTATTTGAGGCGGCGGTGGCGGGCGCCATTCCGGTCATCCGTCCGCTCACCCAG
TCGCTGGCCGGCGACACGGTGCTGCGAGTGGCCGGGATCGTCAACGGCACCACCAACTACATC
15 CTCTCGGCGATGGACAGCACCGGCGCTGACTATGCCAGCGCCCTGGCCGACGCAAGTGCGCT
GGGCTATGCGGAGGCTGATCCCACCGCAGACGTCGAAGGCTACGACGCCCGCGGCAAGGCAG
CGATCCTGGCATCCATTGCCTTCCACACCCGGGTGACCGCAGACGACGTGTATCGCGAAGGCA
TCACCAAGGTCACTCCGGCCGACTTCGGATCCGCGCACGCGCTGGGTTGCACCATCAAAGTGC
TGTCGATCTGTGAGCGCATAACCACCGACGAAGGTTGCGAGCGGGTATCGGCCCGCGTCTATC
20 CGGCCCTGGTACCTCTGTGCGATCCGCTTGCCGCGGTCAACGGCGCGTTCAATGCCGTGGTGG
TCGAGGCCGAGGCCGCGGGCCGGCTGATGTTCTACGGCCAGGGCGCGGGCGGCGCGCCGAC
CGCCTCTGCGGTGACCGGTGACCTAGTGATGGCCGCCCGCAACCGGGTACTCGGCAGCCGCG
GCCCCCGTGAGTCTAAATACGCTCAACTTCCGGTGGCACCAATGGGTTTCATTGAAACGCGCTA
TTACGTCAGCATGAACGTCGCCGACAAGCCGGGCGTCTTGTCGCGGTGGCGGCGGAATTCGCG
25 CAAACGCGAGGTGAGCATCGCCGAGGTGCGCCAGGAGGGCGTTGTGGACGAAGGTGGTGCAC
GGGTGGGAGCCCGAATCGTGGTGGTCACGCACCTCGCCACTGACGCCGCACTCTCGGAAACC
GTTGATGCACTGGACGACTTGGATGTCGTGCAGGGTGTGTCCAGCGTGATACGACTGGAAGGA
ACCGGCTTATGA
- 30 >Rv1323 fadA4 acetyl-CoA C-acetyltransferase (aka thiL) TB.seq 1485860:1487026 MW:40049
>emb|AL123456|MTBH37RV:1485860-1487029, fadA4 SEQ ID NO:50
GTGATTGTTGCTGGCGCGCGTACACCCATCGGCAAGTTGATGGGCTCCCTGAAGGATTCAGCG
CCAGCGAGCTGGGTGCCATCGCCATTAAGGGCGCCCTGGAGAAGGCCAACGTGCCGGCGTCC
TTGGTCGAGTACGTGATCATGGGCCAGGTGTTGACCGCGGGTGCCGGGCAAATGCCCGCACG
35 GCAGGCGGCAGTGGCGGCCGGCATCGGTTGGGATGTCCCTGCGCTGACGATCAACAAGATGT
GCCTGTCCGGCATCGACGCAATCGCGCTGGCTGATCAACTATTGGGCCAGAGAGTTCGACG
TGGTGGTGGCCGGCGGTCAGGAGTCGATGACGAAGGCGCCCCACCTGTTGATGAATAGCCGGT

CGGGTTACAAGTACGGCGACGTTACGGTTTTGGACCACATGGCCTACGACGGTCTGCACGACG
TGTTACCGGATCAGCCGATGGGCGCGCTCACCAGCAACGCAACGACGTCGACATGTTCAACC
GCTCCGAACAGGACGAGTACGCGGCTGCGTCCCACAAAAGGCGGCCGCGGCATGGAAGGAC
GGCGTATTGCGCGACGAGGTGATCCCGGTGAACATCCCGCAGCGCACGGGCGATCCACTGCA
5 GTTCACCGAGGACGAGGGGATCCGCGCCAACACCACCGCCGCGCGCTGGCCGGTCTGAAGC
CGGCGTTCCGTGGCGACGGCACCATCACCGCCGGGTGCGCGTCACAGATCTCCGACGGTGCG
GCCGCGTGGTGGTCATGAACCAGGAAAAGGCCAGGAACTGGGGCTGACCTGGCTAGCCGA
GATCGGCGCCACGGTGTGGTGGCCGGGCGGATTCCACACTGCAATCGCAGCCGGCCAACG
CGATCAACAAGGCGCTGGATCGCGAGGGCATCTCGGTGGACCAGCTCGACGTGGTGGAGATCA
10 ACGAGGCGTTCCGTGCGGTGGCATTGGCCTCGATACGCGAACTCGGGCTGAACCCCCAGATCG
TCAACGTCAACGGTGGTGGCATTGCCGTGGGCATCCCCTCGGCATGTCAGGGACGCGAATCA
CGCTACATGCGGCGCTGCAGTTGGCACGCCGGGGATCGGGCGTCGGGGTTGCCGCATTGTGC
GGGCTGGCGGGCAGGGCGACGCACTGATATTGCGGGCCGGATAG

15 >Rv1389 gmk putative guanylate kinase TB.seq 1564399:1565022 MW:22064

>emb|AL123456|MTBH37RV:1564399-1565025, gmk SEQ ID NO:51

GTGAGCGTCGGCGAGGGACCGGACACCAAGCCCACCGCGCGTGGCCAACCGGCGGCAGTGG
GACGTGTGGTGGTGTCTGTCGGTCTTCCGCGGTGCGCAAATCCACGGTGGTTCGGTGTCTGC
GCGAGCGGATCCCGAATCTGCATTTAGTGTCTCGGCCACGACGCGGGCGCCACGCCCCGGGC
20 GAGGTCGACGGTGTGCGACTACCACTTCATCGACCCACCCGCTTTCAGCAGCTCATCGACCAG
GGTGAGTTGCTGGAATGGGCAGAAATCCACGGCGGCCTGCACCGGTGCGGCACTTTGGCCCA
GCCGGTGCGGGCGGCCGCGGCGACTGGTGTGCCGGTGCTTATCGAGGTTGACCTGGCCGGGG
CCAGGGCGATCAAGAAGACGATGCCCGAGGGTGTACCGTGTTTCTGGCGCCACCTAGCTGGC
AGGATCTTCAGGCCAGACTGATTGGCCGCGGCACCGAAACAGCTGACGTTATCCAACGCCGCC
25 TGGACACCGCGCGGATCGAATTGGCAGCGCAGGGCGACTTTGACAAGGTCGTGGTGAACAGGC
GATTAGAGTCTGCGTGTGCGGAATTGGTATCCTTGCTGGTGGGAACGGCACCGGGCTCCCCGT
GA

>Rv1407 fmu similar to Fmu protein TB.seq 1583099:1584469 MW:48494

30 >emb|AL123456|MTBH37RV:1583099-1584472, fmu SEQ ID NO:52

ATGACCCCTAGATCGCGTGGGCCGCGCCGCGCGCTGGACCCGGCGCGTCGTGCGGCCTT
CGAGACGCTGCGGGCGGTTAGTGCGCGCGACGCCTACGCGAACCTGGTGTGCCCCGCGCTGC
TGGCCCAACGCGGTATCGGCGGTGCGGACGCCGCGTTGCCACCGAGCTGACATACGGCACC
TGCCGAGCCCCGCGGCCTGCTCGACGCGGTATCGGTGCGGCCGCGGAGCGTTGCGCGCAGGC
35 GATCGATCCGGTGTCTGCTAGACCTGTTGCGGCTCGGCACCTACCAATTGCTGCGCACGCGGGT
CGACGCACACGCCGCAGTGTGACCACCGTCGAGCAGGCCGGAATCGAATTGATTGCGCGC
GAGCAGGTTTCGTCAACGGTGTACTACGAACGATCGCCGGCCGAGACGAGCGGTCTGGGTTG

GCGAACTCGCTCCTGATGCGCAGAACGATCCGATCGGGCATGCCGCGTTCTGTGCATGCGCATC
CCCATGGATCGCCAGGCCTTTGCTGACGCGTTGGGCGCGGCGGTGCGGGAGCTCGAGGCA
GTTTTGGCCAGCGACGACGAACGGCCAGCGGTGCACCTGGCGGCACGCCCCGGGGTGCTGAC
CGCCGGCGAACTGGCCCGCGCGGTGCGCGGAACCGTCGGTCGGTATTGCGCGTTTGCGGTGT
5 ATCTGCCGCGCGGTGACCCGGGGCGACTGGCGCCGGTGCAGCGACGGCCAAGCGCTGGTCCA
GGACGAGGGCAGCCAGTTAGTCGCCCCGAGCATTGACCCTGGCGCCAGTCGACGGCGATACCG
GACGGTGGCTGGACCTGTGTGCCGGACCGGGCGGCAAGACCGCGCTGTGGCGGGGCTGGGT
TTGCAGTGCGCAGCCCGGGTGAACCGGTGGAACCCTCGCCACACCGCGCGGACCTGGTAGC
ACAGAACACCCGCGGGCTGCCGGTTGAGCTCTTGCCTGTCGACGGGCGGCACACCGACCTCG
10 ACCCGGGTTTCGACCGGGTGCTGGTGGATGCGCCCTGCACCGGGCTGGGCGCGTTACGCCGT
CGGCCGAGGCCCGTTGGCGTCGTACGCCGGCGGACGTAGCGGCACTGGCCAAGCTACAACG
CGAGTTGTTGAGCGCCGCCATCGCGCTGACTCGGCCCGGCGGTGTCGTGCTCTATGCCACATG
CTCGCCGACCTGGCCGAGACTGTGGGTGCTGTCGCCGACGCGCTACGCCGACATCCGGTTCA
CGCGCTCGATACCCGCCACTGTTGAGCCGGTGATCGCGGGGCTGGGGGAGGGGCCCCACG
15 TTCAGCTGTGGCCGACCGGCACGGTACCGACGCCATGTTGCGCGCGCGTTCGCGCGCCTG
ACGTGA

>Rv1409 ribG riboflavin biosynthesis TB.seq 1585192:1586208 MW:35367

>emb|AL123456|MTBH37RV:1585192-1586211, ribG SEQ ID NO:53

20 ATGAACGTGGAGCAGGTCAAGAGCATCGACGAGGCTATGGGTCTCGCCATCGAGCACTCCTAC
CAGGTCAAAGGCACGACTTATCCAAAACCCCACTGGGGGCCGTATTGTGGATCCCAACGGT
CGGATCGTCGGCGCCGGCGGCACCGAGCCGGCCGGTGGCGATCATGCCGAGGTGGTGGCGC
TGCGCCGGGGCCGGCGGATTGGCTGCCGGCGCCATCGTGGTGGTCACCATGGAACCTGTAAC
CACTACGGCAAGACTCCGCCATGCGTGAACGCTCTGATCGAAGCCAGGGTGGGGACGGTGGTC
25 TACGCCGTCGCCGACCCGAACGGGATCGCTGGGGGTGGCGCGGGCCGGCTGTCAGCAGCGG
GCCTACAGGTGCGGTCCGGGGTGTGGGTGAACAGGTGGCGGCCGGACCGCTGCGGGAGTGG
CTCCACAAGCAACGCACCGGTCTGCCGATGTCACTGGAAGTACGCCACCAGCATCGACGGC
CGCAGCGCCGCCGCCGACGGCTCCAGCCAGTGGATCTCCAGCGAGGCCGCACGCCTGGATCT
GCATCGCCGCCGCGCCATCGCCGACGCGATCTTGGTCGGCACCGGCACCGTCCTCGCCGACG
30 ACCCGGCCCTGACCGCGCGGCTGGCCGACGGCTCGCTGGCGCCGCGAGCAGCCGCTGCGCGT
GGTGGTGGGCAAGCGCGACATACCGCCGGAAGCACGGGTCCTCAACGACGAGGCACGCACCA
TGATGATCCGCACCCACGAACCTATGGAGGTGCTCAGGGCGTTGTGCGATCGCACCGACGTGC
TGCTGGAAGGAGGTCCACCCCTCGCCGGCGCCTTCCTACGAGCGGGTGGGATCAACCGGATCC
TGGCCTACGTGCGACCGATCCTGTTGGGCGGTCCGGTTACCGCGGTGATGACGTGCGGGGTG
35 CCAACATACCAACGCGTTGCGTTGGCAGTTCGACAGCGTCGAAAAGGTGCGACCGGATCTGTT
GCTGAGCTTGGTGGCTCGTTAG

>Rv1440 secG TB.seq 1617715:1618065 MW:12140

>emb|AL123456|MTBH37RV:1617715-1618068, secG SEQ ID NO:54

GTGGCAGGCGTGACAGCCGCGGTCACTGCACGCCTCAAAGCCGATGAGGCGCGACGGCCTGG
GTTCTACGCGGCAGGCAGCGGTCCGCTGCCGCAGGTTCCGGGGAGTACGCTACCCGTCATGG
5 AATTGGCCCTGCAGATCACGCTGATCGTCACGAGCGTGCTGGTGGTGTGTTAGTACTGCTGCA
CCGGGCCAAGGGTGGCGGGCTATCGACACTGTTCCGGCGGTGGTGTGCAGTCAAGCCTGTCCG
GCTCGACGGTGGTGGAGAAGAACCTGGACCGGTTGACGCTGTTGTTACCGGCATCTGGCTGG
TGTCATCATCGGCGTGGCGTTGCTCATCAAATACCGCTAG

10 >Rv1484 inhA TB.seq 1674200:1675006 MW:28529

>emb|AL123456|MTBH37RV:1674200-1675009, inhA SEQ ID NO:55

ATGACAGGACTGCTGGACGGCAAACGGATTCTGGTTAGCGGAATCATCACCGACTCGTCGATCG
CGTTTCACATCGCACGGGTAGCCCAGGAGCAGGGCGCCAGCTGGTGCTACCGGGTTTCGAC
CGGCTGCGGCTGATTCAGCGCATCACCGACCGGCTGCCGGCAAAGGCCCGCTGCTCGAACT
15 CGACGTGCAAAACGAGGAGCACCTGGCCAGCTTGCCGGCCGGGTGACCGAGGCGATCGGGG
CGGGCAACAAGCTCGACGGGGTGGTGCATTGATTGGGTTGATGCCGCAGACCGGGATGGGC
ATCAACCCGTTCTTCGACGCGCCCTACGCGGATGTGTCCAAGGGCATCCACATCTCGGCGTATT
CGTATGCTTCGATGGCCAAGGCGCTGCTGCCGATCATGAACCCCGGAGGTTCCATCGTCGGCA
TGGACTTCGACCCGAGCCGGGCGATGCCGGCCTACAACCTGGATGACGGTCGCCAAGAGCGCG
20 TTGGAGTCGGTCAACAGGTTGCTGGCGCGGAGGCCGGCAAGTACGGTGTGCGTTGCAATCTC
GTTGCCGCAGGCCCTATCCGGACGCTGGCGATGAGTGCGATCGTCGGCGGTGCGCTCGGCGA
GGAGGCCGGCGCCAGATCCAGCTGCTCGAGGAGGGCTGGGATCAGCGCGCTCCGATCGGCT
GGAACATGAAGGATGCGACGCCGGTCGCCAAGACGGTGTGCGCGCTGCTGTCTGACTGGCTG
CCGGCGACCACGGGTGACATCATCTACGCCGACGGCGGCGCGCACACCCAATTGCTCTAG

25

>Rv1617 pykA pyruvate kinase TB.seq 1816187:1817602 MW:50668

>emb|AL123456|MTBH37RV:1816187-1817605, pykA SEQ ID NO:56

GTGACGAGACGCGGGAAAATCGTCTGCACTCTCGGGCCGGCCACCCAGCGGGACGACCTGGT
CAGAGCGCTGGTCGAGGCCGGAATGGACGTCGCCGAATGAACTTCAGCCACGGCGACTACGA
30 CGATCACAAGGTCGCCATGAGCGGGTCCGGGTAGCCTCCGACGCCACCGGGCGCGCGGTG
GCGTGCTCGCCGACCTGCAGGGCCCCGAAGATCAGGTTGGGACGCTTCGCCTCCGGGGCCACC
CACTGGGCGGAAGGCGAAACCGTCCGGATCACCGTGGGCGCCTGCGAGGGCAGCCACGATCG
GGTGTCCACCACCTACAAGCGGCTAGCCCAGGACGCGGTGGCCGGTGACCGGGTGTGGTTCG
ACGACGGCAAAGTCGCATTGGTGGTTCGACGCCGTCGAGGGCGACGACGTGGTCTGCACCGTC
35 GTCGAAGGCGGCCCCGGTCAGCGACAACAAGGGCATCTCGTTGCCCGGAATGAACGTGACCGC
GCCGGCCCTGTGCGAGAAGGACATCGAGGATCTCACGTTGCGGCTGAACCTCGGCGTCGACAT
GGTGGCGCTTTCCTTCGTCCGCTCCCCGGCCGATGTCGAACTGGTCCACGAGGTGATGGATCG

GATCGGGCGACGGGTGCCGGTGATCGCCAAGCTGGAGAAGCCGGAAGCCATCGACAATCTCG
AAGCGATCGTGCTGGCGTTCGACGCCGTGTCATGGTCGCTCGGGGCGACCTAGGTGTTGAGCTGC
CGCTCGAAGAGGTCCCGCTGGTACAGAAGCGAGCCATCCAGATGGCCCCGGGAGAACGCCAAG
CCGGTCATTGTGGCGACCCAGATGCTCGACTCGATGATCGAGAACTCGCGGCCGACCCGAGCT
5 GAGGCCTCCGACGTCGCCAACGCGGTGCTCGATGGCGCCGACGCGCTGATGCTGTCCGGGGA
AACCTCGGTAGGGAAGTACCCCTTGCTGCGGTCCGGACAATGTCGCGCATCATCTGCGCGGT
CGAGGAGAACTCCACGGCCGACCGCCGTTGACACACATTCCCCGGACCAAGCGTGGGGTCAT
CTCGTATGCGGCCCGTGACATCGGCCAAGCGACTCGACGCCAAGGCCTTGTTGGCCTTCACTCA
GTCCGGTGATAACCGTGCGGCGACTGGCCCCGCTGCATACCCGCTGCCGCTGCTGGCCTTAC
10 CGCGTGGCCCCGAGGTGCGCAGCCAACTGGCGATGACCTGGGGCACCAGACGTTTCATCGTGC
CGAAGATGCAGTCCACCGATGGCATGATCCGCCAGGTGACAAATCGCTGCTCGAACTCGCCC
GCTACAAGCGTGGTGACTTGGTGGTCATCGTCGCGGGTGCGCCGCCAGGCACAGTGGGTTCGA
CCAACCTGATCCACGTGCACCGGATCGGGGAAGATGACGTCTAG

15 >Rv1630 rpsA 30S ribosomal protein S1 TB.seq 1833540:1834982 MW:53203
>emb|AL123456|MTBH37RV:1833540-1834985, rpsA SEQ ID NO:57
ATGCCGAGTCCCACCGTCACCTCGCCGCAAGTAGCCGTCAACGACATAGGCTCTAGCGAGGAC
TTTCTCGCCGCAATAGACAAAACGATCAAGTACTTCAACGATGGCGACATCGTCAAGGCACCA
TCGTCAAAGTGGACCGGGACGAGGTGCTCCTCGACATCGGCTACAAGACCGAAGGCGTGATCC
20 CCGCCCCGGAAGTGTCCATCAAGCACGACGTGACCCCAACGAGGTGCTTCCGTGCGTGACG
AGGTGCAAGCCCTGGTGCTCACCAAGGAGGACAAAGAGGGCCGGCTCATCCTCTCCAAGAAAC
GCGCGCAGTACGAGCGTGCTGCGGGCACCATCGAGGCGCTCAAGGAGAAGGACGAGGCCGTC
AAGGGCACGGTCATCGAGGTGCTCAAGGGTGGCCTGATCCTCGACATCGGGCTGCGCGGTTTC
CTGCCCGCCTCGCTGGTGGAGATGCGCCGGGTGCGCGACCTGCAGCCCTACATCGGCAAGGA
25 GATCGAGGCCAAGATCATCGAGCTGGACAAGAACC GCAACAACGTGGTGCTGTCCCGTCGCGC
CTGGCTGGAGCAGACCCAGTCCGAGGTGCGCAGCGAGTTCTGAATAACTTGCAAAAAGGCAC
CATCCGAAAGGGTGTCGTGTCCTCGATCGTCAACTTCGGCGCGTTCTGTCGATCTCGGCGGTGT
GGACGGTCTGGTGATGTCTCCGAGCTATCGTGGAAGCACATCGACCACCCGTCCGAGGTGGT
CCAGGTTGGTGACGAGGTACCGTTCGAGGTGCTCGACGTGACATGGACCGTGAGCGGGTTTC
30 GTTGTCATCAAGGCGACTCAGGAAGACCCGTGGCGGCACTTCGCCCCGCACTCACGCGATCGG
GCAGATCGTGCCGGGCAAGGTCACCAAGTTGGTTCCGTTCCGTGCATTCTGTCGCGTCGAGGA
GGGTATCGAGGGCCTGGTGACATCTCCGAGCTGGCCGAGCGTCACGTGAGGTGCCCGATC
AGGTGGTTGCCGTGCGCGACGACGCGATGGTCAAGGTATCGACATCGACCTGGAGCGCCGTC
GGATCTCGTTGTGCTCAAGCAAGCCAATGAGGACTACACCGAGGAGTTGACCCGGCGAAGT
35 ACGGCATGGCCGACAGTTACGACGAGCAGGGCAACTACATCTTCCCCGAGGGCTTCGATGCCG
AAACCAACGAATGGCTTGAGGGATTGAAAAGCAGCGCGCCGAATGGGAAGCTCGGTACGCCG
AGGCCGAGCGCCGGCACAAGATGCACACCGCGCAGATGGAGAAGTTCGCCGCCGCCGAGGGCG

GCTGGACGCGGCGCGGACGATCAGTCGTCGGCCAGTAGCGCACCGTCGGAAAAGACCGCGGG
TGGATCACTGGCCAGCGACGCCAGCTGGCGGCCCTGCGGGAAAACTCGCCGGCAGCGCTT
GA

5 >Rv1631 - TB.seq 1835011:1836231 MW:44669

>emb|AL123456|MTBH37RV:1835011-1836234, Rv1631 SEQ ID NO:58

ATGCTGCGCATCGGGCTGACCGGCGGCATTGGCGCCGGGAAGTCGTTGCTGTCCACGACGTT
TCGCAATGCGGCGGAATCGTTGTGACGGCGATGTGTTGGCGCGTGAAGTGGTCCAGCCGGGG
ACCGAGGGGCTGGCCTCGCTGGTCGACGCGTTCCGGTCGCGACATCCTGCTTGCAGACGGAGC
10 GCTGGACCGGCAGGCGTTGGCGGCCAAGGCGTTTCGAGATGACGAGTCGCGCGGTGTGCTCA
ACGGAATCGTGACCCGCTGGTCGCCCCGGCGCCGATCCGAGATCATCGCGGCGGTTTCGGGG
GACGCGGTTGTGGTCTGAAGATATCCACTGCTGGTGGAATCCGGGATGGCGCCATTGTTCCGC
TGGTGGTGGTGGTGACGCGGACGTCGAGCTACGGGTGCGACGGCTGGTCGAGCAACGCGGGC
ATGGCCGAAGCCGACGCCCCGGGCTAGGATCGCTGCGCAGGCCAGCGACCAGCAGCGTCGTGC
15 CGTCGCGGACGTCTGGCTGGACAACCTCGGGCAGCCAGAGGATTTGGTGCGGCGGGCCCCGCG
ACGTCTGGAACACGCGCGTCCAGCCCTTCGCGCACAACTGGCCCAACGTCAGATTGCGCGCG
CGCCGGCTAGGTTGGTGCCGGCGGATCCAAGCTGGCCGGATCAGGCGCGGCGCATCGTCAAC
CGGCTAAAGATCGCGTGCGGGCATAAGGCCCTTGCGAGTTGACCACATTGGGTCAACCGCCGTG
TCGGGCTTCCCCGATTTCTAGCCAAGGATGTCATCGACATCCAGGTCACCGTCGAATCACTTG
20 ACGTGGCCGACGAGCTGGCCGAGCCCTTGCTGGCCGCGGGCTACCCACGCCTCGAGCACATC
ACCCAGGACACCGAAAAGACCGACGCTCGCAGCACCGTCGGCCGCTACGACCACACCGACAGT
GCCGCTCTGTGGCACAAGCGCGTGACGCCTCGGCGGATCCCGGTGCGCCGACCAACGTGCA
CCTGCGGGTGACGGCTGGCCCCAACCAACAGTTCGCCCTGCTGTTCTGTCGACTGGCTGGCGGG
CAATCCCGGCGCGAGAGAAGACTATTTGACGGTCAAGTGTGACGCCGACAGGCGCGCCGACG
25 GTGAGCTCGCGCGCTACGTCACCGCCAAGGAGCCGTGGTTCCTGGATGCCTACCAGCGGGCAT
GGGAGTGGGCGGATGCGGTGCACTGGCGTCCCTGA

>Rv1706c - TB.seq 1932695:1933876 MW:39779

>emb|AL123456|MTBH37RV:c1933876-1932692, PPE SEQ ID NO:59

30 ATGACCCTCGATGTCCCGGTCAACCAGGGGCATGTCCCCCGGGCAGCGTCGCCTGCTGCCTT
GTTGGGGTCACCGCCGTTGCTGACGGCATCGCCGGGCATTCCCTGTCCAACTTTGGGGCGTTA
CCTCCCGAGATCAATTCGGGTGCTATGTATAGCGGTCCGGGATCCGGGCCACTGATGGCTGCC
GCGGCGGCCTGGGACGGGCTGGCCGAGAGTTGTCGTCGGCAGCGACTGGCTACGGTGCGG
CGATCTCGGAGCTGACAAACATGCGGTGGTGGTCGGGGCCGGCATCGGATTGATGGTGGCC
35 GCCGTCTGCCCTTTGTCGGCTGGCTGAGTACCACGCGACGCTAGCCGAACAGGCCGCGATG
CAGGCTAGGGCGGGCCGACGCGCCTTTGAAGCCGCCTTCGCCATGACGGTGCCCCCGCGGGC
GATCGCGGCCAACCGGACCTTGTTGATGACGCTCGTCGATACCAACTGGTTCGGGCAAAACAC

GCCGGCGATCGCCACCACCGAGTCCCAATACGCCGAGATGTGGGCCCAAGACGCCGCCGCGA
TGTACGGCTATGCCAGCGCCGCGGCACCCGCCACGGTTTTGACTCCGTTGACACCACCGCCGC
AAACCACCAACGCGACCGGCCTCGTCGCCACGCAACAGCGGTGGCCGCGCTGCGGGGGCAG
CACAGCTGGGCCGCGGCGATTCCATGGAGCGACATACAGAAATACTGGATGATGTTCTCTGGGC
5 GCCCTCGCCACTGCCGAAGGGTTTACATTTACGACAGCGGTGGGTTAACGCTGAATGCTCTGCAGT
TCGTGCGCGGGATGTTGTGGAGCACCGCATTGGCAGAAGCCGGTGCGGCCGAGGCAGCGGCC
GGCGCGGGTGGAGCCGCTGGATGGTCGGCGTGGTCGCAGCTGGGAGCTGGACCGGTGGCGG
CGAGCGCGACTCTGGCCGCCAAGATCGGACCGATGTCGGTGCCGCCGGGCTGGTCCGCACCG
CCCCGCCACGCCCCAGGCGCAAACCGTCGCGCGATCGATTCCCGGTATTGCGAGCGCCGCCGA
10 GGCGGCTGAAACATCGGTCCTACTCCGGGGGGCACCAGTCCGGGCAGGAGTCGCGCCGCC
ATATGGGACGCCGATATGGAAGACGACTCACCGTGATGGCTGACCGGCCGAACGTCCGATAG

>Rv1745c - similar to Q46822 ORF_O182 TB.seq 1971381:1971989 MW:22490

>emb|AL123456|MTBH37RV:c1971989-1971378, Rv1745c SEQ ID NO:60

15 ATGACCCGCAGCTACCGGCCAGCTCCACCGATCGAGCGGGTGGTTTTGCTCAACGACCGCGGC
GACGCGACAGGTGTGGCCGACAAGGCCACCGTGACACCGGCGACACCCCTTTGCACCTCGC
GTTCTCCAGCTATGTGTTGATCTGCACGATCAGCTGTTGATCACGCGGCGGGCCGCCACCAAG
AGGACGTGGCCGGCGGTATGGACCAACAGTTGCTGCGGGCACCCCTGCCTGGCGAATCGCT
ACCCGGCGCCATACGCCGGCGGCTCGCTGCCGAACTCGGACTGACCCAGATCGGGTCGATC
20 TGATCCTGCCGGGGTTCGGCTACCGGGCCGCTATGGCCGATGGCACCGTGGAACGAGATCT
GCCCCGTCTACCGAGTCCAGGTTGACCAACAGCCCCGGCCGAACTCGGACGAGGTGACGCG
ATCCGCTGGTTGTCTGGGAACAATTCGTGCGCGATGTTACCGCCGGCGTAATCGCCCCGGTAT
CCCCTTGGTGCCGCTCACAACCTGGGCTACCTGACCAAACCTTGGACCATGTCCGGCACAGTGGC
CCGTGGCCGACGACTGCCGGCTACCGAAAGCCGCACATGGTAATTAA

25

>Rv1800 - TB.seq 2039451:2041415 MW:67068

>emb|AL123456|MTBH37RV:2039451-2041418, PPE SEQ ID NO:61

ATGCTGCCGAATTTGCGGGTGTGCCCCCGAGGTCAATTCGGCGAGGGTGTTCGCCGGTGCG
GGGTGCGCGCCGATGTTAGCGGCAGCGGCCCTGGGATGATCTAGCCTCCGAGCTGCATTGT
30 GCTGCAATGTCAATTCGGGTGCGTTACGTGCGGATTGGTGGTTGGTGGTGGCAGGGATCGGCC
TCGGCGGCGATGGTGGACGCAGCCGCGTGTACATCGGGTGGCTGAGCACGTGCGGCTGCCCA
CGCCGAGGGCGCGGCCGGTCTGGCTCGGGCCGCGGTATCGGTGTTGAGGAGGCGCTGGCC
GCGACGGTGCATCCGGCGATGGTTGCGGCAAATCGCGCCAGGTGGCGTCGCTGGTAGCGTC
GAACTTGTGTTGGGCAGAACGCGCCTGCGATCGCCGCGCTCGAATCCTTGATGAGTGTATGTGG
35 GCCCAGGATGCAGCGGCCATGGCGGGTTATTACGTTGGGGCTTCGGCGGTGGCCACACAGTTG
GCATCGTGGCTGCAACGGCTACAGAGCATCCCCGGCGCCGAGTCTTGATGCCCGTCTGCCG
AGCTCGGCCGAGGCACCGATGGGAGTCGTCCGCGCGGTCAACAGCGCGATCGCCGCCAATGC

GGCTGCGGCACAAACCGTTGGCCTGGTCATGGGAGGCAGCGGCACGCCAATACCGTCGGCCA
GATATGTCGAGCTCGCGAACGCGCTGTACATGAGTGGCAGCGTCCCGGGTGTATCGCGCAGG
CGCTCTTCACGCCCCAAGGGCTCTACCCGGTGGTCGTGATCAAGAACCTCACTTTCGATTCTC
GGTGGCGCAGGGTGCCGTCACTCTCGAAAGTGCATTTCGGCAGCAAATTGCCGCCGGCAACAA
5 CGTCACCGTCTTCGGCTACTCGCAGAGCGCCACGATCTCGTCACTAGTGATGGCCAATCTTGCG
GCTTCGGCCGACCCGCCGTCTCCAGACGAGCTTTCCTTCACGCTGATCGGCAATCCCAACAACC
CCAATGGCGGGGTGGCCACCAGGTTCCCGGGGATCTCCTTCCAAGCTTGGGCGTGACGGCCA
CCGGGGCCACTCCGCACAATCTGTACCCGACCAAGATCTACACCATCGAATACGACGGCGTCG
CCGACTTTCGCGGTACCCGCTCAACTTTGTGTGACCCCTCAACGCCATTGCCGGCACCTACTA
10 CGTGCACTCCAATACTTCATCCTGACGCCGGAACAAATTGACGCAGCGGTTCCGCTGACCAAT
ACGGTCGGTCCACGATGACCCAGTACTACATCATTCGCACGGAGAACCTGCCGCTGCTAGAG
CCACTGCGATCGGTGCCGATCGTGGGGAACCCACTGGCGAACCTGGTTCAACCAAACCTTGAAG
GTGATTGTAACTGGGCTACGGCGACCCGGCCTATGGTTATTGACCTCGCCGCCCAATGTTG
CGACTCCGTTCCGGTTGTTCCAGAGGTGAGCCCGTCTCGTATCGCCGACGCTCTCGTCGCCG
15 GGACCCAGCAGGGAATCGGCGATTTGCCTACGACGTGAGCCACCTCGAACTGCCGTTGCCGG
CAGACGGGTCGACGATGCCAAGCACCGCACCCGGGCTCGGGTACGCCGGTCCCCCGCTCTCG
ATCGACAGCCTGATAGACGACCTGCAGGTGGCTAACCGCAACCTCGCCAACACGATTTGAAG
GTGGCCGCGACGAGCTACGCGACGGTGCTCCCAACCGCCGACATCGCCAATGCGGCCTTGAC
GATCGTGCCGTCGTACAACATCCACCTTTTTTTGGAGGGCATCCAGCAAGCGCTCAAGGGCGAC
20 CCGATGGGACTCGTCAACGCGGTGCGGATACCCACTCGCGGCCGACGTGGCACTGTTACGGCC
GCAGGCGGTCTTCAGCTCTTGATCATCATCAGCGCGGGCCGAACGATTGCCAATGACATCTCGG
CCATTGTCCCCTGA

>Rv1844c gnd 6-phosphogluconate dehydrogenase (Gram -) TB.seq 2093732:2095186
25 MW:51548 >emb|AL123456|MTBH37RV:c2095186-2093729, gnd SEQ ID NO:62
ATGAGTTCGTCGGAATCGCCAGCCGGCATCGCGCAGATCGGCGTCACTGGCCTGGCCGTGATG
GGTTCCAACATCGCCCGAACTTCGCCCGGCACGGCTACACCGTGGCAGTGACAATCGGTGCG
GTGCGCAAGACCGACGCGCTGCTTAAGGAGCACAGCTCAGACGGCAAGTTCGTGCGCAGTGAA
ACGATCCCCGAATTTCTTGCCGCACTGGAAAAACCGCGTCGGGTGCTGATCATGGTCAAGGCC
30 GGAGAGGCCACTGACGCTGACGCTGTATCAACGAACTTGCTGACGCCATGGAACCCGGCGAC
ATCATCATCGACGGCGGCAATGCGTTGTACACCGACACCATGCGCCGCGAGAAAGCGATGCGT
GAGCGGGGCTTGCACTTCGTGCGGGCCGGGATCTCCGGCGGCGAAGAGGGCGCGTTGAACGG
GCCGTCGATCATGCCCGGCGGACCCGCCGAGTCATACCAATCGCTGGGTCCGCTGCTCGAGGA
GATCTCCGCGCATGTCGACGGCGTGCCGTGCTGCACCCACATTGGCCCGGACGGCTCCGGGC
35 ACTTCGTCAAGATGGTCCACAACGGCATCGAGTACTCCGACATGCAGCTCATCGGTGAGGCCTA
CCAGCTGATGCGCGACGGGCTAGGTCTGACCGCGCCGGCGATCGCCGATGTGTTACCGAGT
GGAACAATGGCGATCTGGACAGCTACCTGGTCGAGATCACCGCCGAGGTGCTGCGGCAGACCG

ATGCCAAGACCGGCAAACCGCTCGTCGACGTCATCGTGGACCGGGCCGAGCAGAAAGGCACC
GGCCGTTGGACCGTCAAGTCCGCGCTGGACCTGGGTGTGCCGGTGACCGGCATCGCCGAAGC
GGTGTTCGCCGCGCTCTCTCGGGATCCGTGGGGCAACGCTCGGCCGCCAGCGGTCTGGCTTC
GGGCAAGCTCGGCGAGCAGCCCGCCGACCCGCCACGTTACCGAAGACGTCCGCCAGGCGT
5 TGTACGCCTCCAAGATCGTGGCCTACGCTCAGGGCTTCAACCAGATCCAGGCCGGCAGCGCCG
AATTCGGCTGGGACATCACGCCGGGCGACCTGGCCACCATCTGGCGTGGCGGCTGCATCATCC
GGGCGAAGTTCCTCAACCACATCAAGGAAGCCTTTGACGCCAGCCCGAACCTGGCCAGTCTGA
TTGTGGCCCCGTATTTCCGCGGCGCCGTGGAATCGGCGATCGACAGTTGGCGGCGTGTGGTGT
CGACGGCGGGCCCAACTGGGTATCCCGACCCCGGGATTCTCGTCGGCCCTGTCGTATTACGACG
10 CGCTGCGCACCGCGCGGCTGCCCGCTGCACTACCCAGGCCAGCGCGACTTCTTCGGCGCA
CACACCTACGGCCGGATCGACGAACCAGGCAAGTTCACACACTATGGAGTTCAGACCGCACC
GAAGTACCGGTGTAG

>Rv1900c lipJ TB.seq 2146246:2147631 MW:49685

15 >emb|AL123456|MTBH37RV:c2147631-2146243, lipJ SEQ ID NO:63
GTGGCGCAGGCTCCCCACATTCACAGGACCCGCTACGCAAAATGCCGCGACATGGATATCGCC
TACCAGGTGCTGGGTGACGGTCCGACGGATCTGCTGGTGTTCGCCGGGGCCGTTCTGTGCCGATC
GACTCGATCGACGACGAGCCATCGCTGTACCGTTTCCATCGCCGTCTTGCGTCATTACGAGGG
TGATCCGCCTCGACCATCGTGGGGTCGGCCTGTCGTCACGGCTCGCCGCGATAACCACGCTGG
20 GGCCGAAGTTCGCGCCAGTTCACGCCATGAACGGACTTGTTCTCGCCGCCGACTACCCCGAGC
ACAATTTTCGCGCCAGTTCACGCCATGAACGGACTTGTTCTCGCCGCCGACTACCCCGAGC
GGGTGCGCAGCCTGATCGTCGTCAACGGCTCGGCGCGCCCACTATGGGCGCCCGACTACCCG
GTAGGCGCCAGGTTCTGTCGAGCTGACCCGTTCTGACGGTGGCGCTGGAACCGGATGCCGTC
GAGCGGGGCTTCGACGTGCTGAGCATCGTGGCTCCTACCGTGGCCGGAGATGACGTGTTTCGA
25 GCCTGGTGGGATCTCGCCGGCAACCGTGCCGGACCGCCGAGCATTGCCCGTGCCGTTTCAAAG
GTCATAGCCGAGGCCGACGTACGAGATGTCTTGGGACACATCGAGGCTCCAACACTGATCTTGC
ACCGTGTCCGATCGACGTACATCCCGGTGGGACATGGTCGCTACCTCGCCGAGCACATCGCTG
GATCCCGCTTGGTCAACTACCCGGCACCGATACCCTGTACTGGGTTGGCGACACCGGGCCGA
TGCTCGATGAAATCGAGGAATTCATCACCGGCGTGCGCGGCGGCGCTGACGCCGAGCGCATGC
30 TTGCCACCATCATGTTTACCGACATCGTCGGCTCGACCCAGCACGCCGCCGCGCTCGGCGACG
ACCGATGGCGCGACCTGTTGGACAACCACGACACCATCGTGTGCCACGAAATCCAGCGGTTTCG
GCGGTGCGGAAGTGAACACGGCCGGTGACGGTTTCGTGCGGACGTTACCCAGTCCGAGTGCC
GCGATCGCGTGCGCGGACGACATCGTCGACGCGGTGCGCGCGCTGGGTATTGAGGTCCGAT
CGGTATTCATGCGGGCGAGGTGAGGTGCGCGATGCCTCGCACGGTACCGACGTGCGCCGGCG
35 TGGCCGTGCATATCGGTGCGCGCGTCTGCGCGCTGGCCGGACCCAGTGAGGTGCTGGTGTCC
TCGACCGTGCGAGACATCGTCGCCGGATCACGGCACCGGTTCCGCCGAGCGTGGTGAGCAGGA

ACTCAAGGGCGTACCGGGCAGATGGCGGCTATGCGTGCTCATGCGCGACGACGCCACCCGCA
CGCGCTAA

>Rv1967 - TB.seq 2210599:2211624 MW:36516

5 >emb|AL123456|MTBH37RV:2210599-2211627, Rv1967 SEQ ID NO:64

ATGAGGGAGAACCTGGGGGGCGTCGTGGTGCGCCTCGGCGTCTTCCTGGCGGTATGCCTGCT
GACGGCGTTTCCTGCTGATTGCCGTCTTCGGGGAGGTGCGCTTCGGCGACGBCAAGACCTACTA
CGCCGAGTTCCGCCAACGTGTCCAATCTGCGAACGGGCAAGCTGGTGCGCATCGCCGGCGTCGA
GGTCGGCAAGGTCACCAGGATCTCCATCAACCCCGACGCGACGGTGCGGGTGCAAGTTCACCGC
10 CGACAACCTCGGTCAACCTCACGCGGGGGCACCCGGGCGGTGATCCGCTACGACAACCTGTTCCGG
TGACCGCTATTTGGCGCTGGAGGAAGGGGGCCGGCGGACTCGCCGTTCTTCGTCCCGGTACAC
GATTCCGTTGGCGCGCACCCAACCGGCGTTGGATCTGGATGCCCTGATCGGTGGATTCAAGCC
GCTGTTTCGTGCGCTGAACCCCGAGCAGGTCAACGCGCTGAGCGAACAGTTGCTGCACGCGTT
TGCCGGACAGGGGGCCACGATCGGGTCATTGCTGGCCCGAGTCCGCGGCGGTGACCAACACCC
15 TGGCCGACCGTGATCGGCTGATCGGGCAGGTGATCACCAACCTCAACGTGGTGCTGGGCTCGC
TGGGCGCTCACACCGATCGGTTGGACCAGGCGGTGACGTGCTATCAGCGTTGATTACCGGC
TCGCGCAACGCAAGACCGACATCTCCAACGCCGTGGCCTACACCAACGCCGCCGCCGGCTCG
GTCGCCGATCTGCTGTGCGCAGGCTCGCGCGCCGTTGGCGAAGGTGGTTCGCGAGACCGATCG
GGTGGCCGGCATCGCGGCCGCCGACCAGACTACCTCGACAATCTGCTCAACACGCTGCCGGA
20 CAAATACCAGGCGCTGGTCCGCCAGGGTATGTACGGCGACTTCTTCGCCTTCTACCTGTGCGAC
GTCGTGCTCAAGGTCAACGGCAAGGGCGGCCAGCCGGTGTACATCAAGCTGGCCGGTCAGGA
CAGCGGGCGGTGCGCGCCGAAATGA

>Rv1975 - TB.seq 2218050:2218712 MW:23650

25 >emb|AL123456|MTBH37RV:2218050-2218715, Rv1975 SEQ ID NO:65

ATGTCGCGTCGAGCATCGGCCACGTGTGCCTTGTCCGCGACCACCGCCGTCGCCATAATGGCT
GCTCCCGCCGCACGGGCCGACGACAAGCGGCTCAACGACGGCGTGGTCGCCAACGTCTACAC
CGTTCAACGTCAGGCCGGCTGCACCAACGACGTCACGATCAACCCGCAACTACAATTGGCCGC
CCAATGGCACACCTCGATCTGCTGAACAACCGGCACCTCAACGACGACACCGGTTCTGACGG
30 ATCCACACCGCAAGACCGCGCGCATGCCGCCGGCTTCGCGGGAAAGTCGCTGAAACCGTGG
CGATCAATCCCGCCGTAGCGATCAGCGGCATCGAGTTGATAAACCAGTGGTACTACAACCCCGC
GTTTTTCGCGATCATGTCCGACTGCGCCAACACCCAGATCGGGGTGTGGTCAGAAAACAGCCC
GGATCGCACCGTCGTGGTGGCCGTTTACGGACAGCCCGATCGACCTTCGCGATGCCGCCAG
GGGAGCGGTAACCGGACCGCCGTCCCGGTGGCCGCGCAAGAGAACGTTCCCTATCGACCCCA
35 GCCCGACTACGACGCCAGCGACGAGATCGAATACGGCATCAACTGGCTGCCATGGATCCTGC
GCGGCGTGTACCCGCCGCCGCCAATGCCGCCGCAAGTAG

>Rv1981c nrdF ribonucleotide reductase small subunit TB.seq 2224221:2225186 MW:36591

>emb|AL123456|MTBH37RV:c2225186-2224218, nrdF SEQ ID NO:66

ATGACCGGCAAGCTCGTTGAGCGGGTGACGCAATCAATTGGAACCGGTTGCTCGATGCTAAA
GATTTGCAGGTCTGGGAACGTTTGACCGGTAACTTTGGTTGCCGGAAGATTCCGCTCTCCA
5 ACGACCTGGCATCTTGCAAACGTTGAGTTCCACCGAGCAGCAGACGACGATCCGGGTGTTCA
CCGGCTTGACCCTGCTCGACACCGCGCAGGCGACGGTGGGAGCAGTGGCCATGATCGACGAC
GCGGTACACCCCCACGAAGAGGCGGTCTTGACCAACATGGCGTTCATGGAGTCAGTGCACGCC
AAGAGCTACAGCTCGATCTTCTCGACCCTGTGCTCGACCAAGCAGATCGACGATGCCTTCGACT
GGTCGGAACAGAACCCTTACCTGCAGCGAAAAGCGCAGATCATCGTCGACTACTACCGCGGTG
10 ACGACGCGCTCAAGCGCAAAGCATCGTCGGTAATGCTGGAGTCCTTCCTGTTCTACTCCGGCTT
CTACCTGCCCATGTACTGGTCGTCGCGGGTAAGCTCACCAACACCGCCGATCTGATCCGGCT
GATCATCCGAGATGAAGCCGTCCACGGCTACTACATCGGCTACAAATGTCAACGAGGTTTGGCC
GACCTGACCGACGCCGAGCGGGCCGACCACCGCGAATACACCTGCGAGCTGCTGCACACGCT
CTACGCGAACGAGATCGACTATGCGCACGACTTGTACGACGAGTTGGGCTGGACCGACGACGT
15 TTTGCCCTACATGCGTTACAACGCCAACAAGGCGCTAGCCAACCTGGGATACCAGCCTGCATTG
GATCGTGACACCTGCCAGGTGAACCCGGCGGTGCGCGCAGCTCTCGACCCCGGTGCAGGGGA
GAACCACGACTTTTTCTCCGGCTCCGGAAGCTCATACGTAATGGGCACCCACCAACCCACCACC
GACACCGACTGGGACTTCTAA

20 >Rv2092c helY helicase, Ski2 subfamily TB.seq 2349335:2352052 MW:99576

>emb|AL123456|MTBH37RV:c2352052-2349332, helY SEQ ID NO:67

GTGACTGAGCTGGCCGAGCTGGACCGGTTACCGCGGAACCTACCGTTCTCGCTCGACGACTTT
CAGCAGCGGGCTTGCAGCGCGCTGGAACGCGGGCCACGGTGTGCTGGTGTGCGCGCCGACCG
GCGCTGGCAAGACGCTGGTCGGCGAGTTCGCCGTGCACCTGGCGCTGGCGGGCCGGCAGTAA
25 TGTTTCTACACCACGCCGCTGAAAGCCCTGAGCAACCAAAAGCACACCGATCTCACAGCACGCT
ACGGCCGTGACCAGATCGGGCTGCTGACCGGTGACCTGTCCGTCAACGGCAACGCGCCGGTG
GTGGTGATGACCACCGAAGTGCTGCGCAACATGCTCTACGCGGATTGCGCTGCGCTGCAGGGG
CTTTCCTATGTGGTGATGGATGAGGTGCATTTCTCGCCGACCGGATGCGGGGTCCGGTGTGG
GAGGAGGTGATCCTGCAACTGCCCCGACGACGTGCGGGTGGTCAGCCTGTGCGCGACGGTGAG
30 CAACGCCGAGGAGTTCGGCGGTTGGATCCAGACGGTGCAGGGGCGACACCACGGTGGTGGTCG
ACGAGCATCGGCCGGTGCCGTTGTGGCAACACGTCTTGGTGGGCAAGCGCATGTTGACCTGT
TCGATTACCGGATCGGCGAAGCCGAAGGGCAGCCCCAAGTCAACCGCGAGTTGCTGCGCCACA
TCGCGCATCGCCGTGAGGCCGACCGGATGGCCGATTGGCAGCCTCGGCGCCGAGGCTCGGGC
CGGCCCGGCTTCTACCGGCCACCCGGCCGACCCGAGGTGATCGCCAAACTCGACGCTGAAGG
35 GCTGTTGCCGGCGATCACCTTCGTGTTCTCCCGGGCCGGTTGTGACGCCGCGGTCAACCAATG
CCTGCGGTACCGCTGCGGTTGACCAGCGAAGAGGAGCGCGCACGGATCGCCGAGGTGATCG
ACCACCGCTGCGGTGACCTGGCCGACTCCGACCTGGCGGTACTCGGCTACTACGAATGGCGG

GAAGGGTACTGCGCGGTCTGGCCGCCACACGCGGGCATGTTGCCGGCCTTCCGGCACAC
GGTGGAGGAGCTGTTACCGCCGGTTTGGTCAAGGCTGTATCGCCACCGAGACTCTGGCGCT
CGGTATCAACATGCCGGCCCGCACGGTGGTGTGGAGCGGCTGGTGAAGTTCAACGGTGAGCA
GCACATGCCGCTGACGCCGGGGGAGTACACCAACTGACCGGTGCGCGCCGGCCGGCGCGGTA
5 TCGACGTCGAGGGTCACGCGGTGGTGATCTGGCACCCGGAATTGAACCGTCCGAGGTGGCG
GGCCTGGCCTCCACCCGCACCTTTCCGCTGCGCAGCTCGTTTGCCCCGTCGTACAACATGACG
ATCAACCTGGTGCACCGGATGGGTCCGCAACAGGCGCACCGACTGCTCGAGCAGTCGTTCCGCC
CAATATCAGGCCGACCGATCCGTGGTGGACTGGTCCGCGGAATTGAGCGGGGCAACAGGATA
CTCGGCGAGATCGCAGCCGAAGTGGGCGGATCTGATGCGCCCATCCTCGAATACGCTCGATTG
10 CGCGCGCGGGTGTCCGAGCTGGAACGTGCGCAGGCCCGCGCGTTCGCGGTTACAGCGACGGC
AGGCGGCCACCGATGCGCTGGCCGCGCTGCGCCGCGGTGACATCATCACCATCACCCACGGC
CGCCGCGGTGGTCTGGCCGTCGTCTGGAATCAGCCCGCGACCGCGACGACCCGCGTCCGCT
GGTGCTAACCGAACACCGATGGGCGGGACGGATCTCCTCGGCCGACTACTCGGGCACGACGC
CGGTGGGGTCGATGACGCTGCCAAGCGGGTGGAGCACCGCCAGCCGCGGGTCCGGCGTGA
15 CCTGGCCTCGGCGCTGCGATCGGCAGCCGCGGGTCTGGTTATTCCAGCCGCCCGGCGCGTCA
GCGAGGCCGGCGGGTTTACGATCCGGAGCTGGAGTCGTGCGCGGAACAATTGCGCCGTCAT
CCGGTGCATACCTCGCCCGGGCTCGAGGACCAGATCCGCCAGGCCGAGCGTTACTTACGCATC
GAACGCGACAACGCGCAATTAGAGAGGAAGGTGCGCCGCCACCAACTCGTTGGCCCGCAC
GTTTCGACCGATTCTGTCGGGCTGCTACCGAACGGGAGTTTCATCGATGGCCCGGCCACTGATCC
20 CGTGGTCACCGACGACGGCCGGCTGCTGGCGCGGATTTACAGCGAGAGCGACCTGTTGGTGG
CCGAGTGCCTACGTACAGGTGCGTGGGAGGGTTTAAAGCCGGCCGAATTGGCGGGGGTGGTG
TCGGCGGTGGTCTACGAGACGCGCGGTGGTGACGGCCAGGGCGCCCCGTTCCGAGCCGATGT
GCCACACCGCGGTTACGGCAGGCTCTGACTCAGACATCAAGGCTGTCCACGACATTGCGCGC
CGACGAGCAGGCACACCGCATCACCCGAGTCGCGAACCCGACGATGGCTTTGTCAGAGTCAT
25 CTACCGCTGGTCGCGAACCGGTGATCTAGCGGCGGCATTGGCCGCTGCCGACGTGAACGGCA
GCGGATACCGTTATTGGCAGGGGATTTCGTGCGTTGGTGCCGTCAGGTGCTCGATCTGCTGG
ACCAAGTTCGTAACGCTGCGCCCAACCCCGAACTGCGGGCTACCGCAAAGCGCGCTATCGGTG
ACATTGGCGCGGGCGTCTGCGCGTTGACGCCGGGTAG

30 >Rv2101 helZ helicase, Snf2/Rad54 family TB.seq 2360238:2363276 MW:111632

>emb|AL123456|MTBH37RV:2360238-2363279, helZ SEQ ID NO:68

ATGCTGGTTTTGCACGGCTTCTGGTCCAACCTCCGGCGGGATGCGGCTGTGGGCGGAGGACTCC
GATCTGCTGGTGAAGAGCCCGAGTCAGGCGCTGCGCTCCGCGCGGCCACACCCGTTCCGCGC
GCCCCGCTGACCTGATCGCCGGCATACTCCGGGCAAACCCGCAACCGCCGTTTTGCTGTTGCC
35 GTCGTTGCGATCGGCGCCGCTGGACTCGCCGGAGCTGATCCGGCTCGCCCCGCGCCCGGCCG
CGCGAACCGATCCGATGCTGTTGGCGTGGACGGTACCGGTGGTGGACCTGGACCCCAACCGCG
GCGTTGGCCGCCTTCGACCAGCCCGCCCCCGACGTCCGCTACGGCGCGTCCGTGACTACCT

GGCCGAGCTGGCCGTTTTTCGCGCGCGAGTTGGTCGAGCGTGGTCGCGTGCTGCCCCAGCTGC
GCCGCGACACCCACGGCGCGGCCGCTGCTGGCGTCCGGTGTTGCAGGGACGCGACGTGGTC
GCGATGACCTCGCTGGTCTCGGCGATGCCGCCGGTCTGCCGCGCCGAAGTTGGTGGGCACGA
CCCCGACGAAGTGGCAACCTCGGCTCTGGACGCGATGGTCGACGCCGCCGTGCGCGCGGCGC
5 TGTCACCGATGGACCTGCTGCCCCCGCGACGGGGTCGCTCCAAACGGCATCGGGCCGTGGAG
GCTTGGCTGACCGCGTTGACCTGCCCCGACGGCCGGTTCGACGCGGAGCCCCAGCAACTCGA
CGCGCTGGCCGAGGCGTTGCGGGCATGGGACGACGTGGTATCGGCACCGTCGGCCCCGGCGC
GGGCGACGTTTTCGGCTGTCCGAAGTCGAGACCGAAAACGAGGAGACGCCCGCGGGCTCGTTG
TGGAGGCTGGAGTTCTTATTGCAGTCGACGCAGGACCCAGCCTGCTGGTCCCCGCCGAGCAG
10 GCATGGAACGACGACGGCAGCCTGCGCCGCTGGCTGGACCGGCCGAGGAGCTGCTGCTGAC
CGAACTGGGCCGGGCCTCTCGGATTTTCCCCGAGCTCGTCCCGGCGCTGCGCACCGCGTGCC
CGTCCGGGCTTGAGCTCGACGCCGACGGCGCCTACCGATTCTGTGGGTACGGCCGCGGTG
CTCGACGAGGCTGGGTTTGGCGTGCTGCTGCCGTCTGGTGGGACCGCCGCCGCAAGCTGGG
CTTGGTCTGTCCGCATATACCCCGGTGACGGCGTGGTGGGCAAGGCCAGCAAGTTCGGCCG
15 CGAGCAGCTCGTCGAGTTCCGCTGGGAGCTGGCCGTGGGCGACGATCCGCTCAGCGAGGAGG
AGATCGCGGCGCTGACCGAAACCAAGTCCCCGCTGATCCGGCTGCGTGGCCAGTGGGTGCGC
CTCGATACCGAACAGATGCGCCGCGGGCTGGAGTTTTTGGAGCGTAAGCCAACCGGCCGCAAG
ACCACCGCCGAGATCCTCGCGCTGGCCGCCAGCCACCCGACGACGTGGACACCCCGCTCGA
GGTCACCGCCGTACGCGCCGACGGCTGGCTCGGGGACCTGCTCGCCGGGGCCGCCGCGGCG
20 TCGCTGCAGCCGTTGGACCCGCCCGACGGATTACCGCGACGCTGCGTCCCTACCAGCAGCGC
GGTCTGGCGTGCTGGCGTTTTTGTCTCGCTCGGTTTGGGACGCTGCCTGGCCGACGACATG
GGCCTGGGCAAGACGGTGCAGCTATTGGCCCTGGAAACCTTGAATCCGTTGAGCGCCACCAG
GATCGCGGCGCTCGGACCCACACTGCTACTGTGCCGATGTGCTTGGTGGGCAACTGGCCGAG
GAAGCGGCCAGGTTTGCACCCAACTGCGGGGTGTACGCCCACCACGGGGGCGCCCGGCTGCA
25 CGGCGAGGCGTTGCGCGACCACCTCGAGCGCACCGACCTGGTCTGAGCACCTATACCACCG
CCACCCGCGACATCGACGAGCTGGCGGAATACGAATGGAACCGGGTGGTGTGGACGAGGCC
CAGGCGGTGAAGAACAGCCTGTCCCGGGCGGCCAAGGCGGTGCGACGGCTACGCGCGGCGC
ACCGGGTCTGCGCTGACCGGGACACCGATGGAGAACCGGCTCGCCGAGCTGTGGTCGATCATG
GACTTCCTCAACCCGGGCTGCTCGGATCCTCCGAACGCTTCCGCACCCGCTACGCGATCCCG
30 ATCGAGCGGCACGGGCACACCGAACCGGCCGAACGGCTGCGCGCATCGACGCGGCCCTACAT
CCTGCGCCGGCTCAAGACCGACCCGGCGATCATCGACGATCTGCCGGAGAAGATCGAGATCAA
GCAGTACTGCCAACTCACCACCGAGCAGGCGTCGCTGTATCAGGCCGTGTCGCCGACATGAT
GGAAAAGATCGAAAACACCGAAGGGATCGAGCGGCGCGGCAACGTGCTGGCCGCGATGGCCA
AGCTCAAACAGGTGTGCAACCACCCCGCCAGCTGCTGCACGATCGCTCCCCGGTCGGTCGGC
35 GGTCCGGGAAGGTGATCCGGCTCGAGGAGATCCTGGAAGAGATCCTGGCCGAGGGCGACCGG
GTGCTGTGTTTTACCCAGTTCACCGAGTTCGCCGAGCTGCTGGTGCCGCACCTGGCCGCACGC
TTCGGCCGTGCCGCCCGAGACATTGCCTACCTGCACGGTGGCACCCCGAGGAAGCGGCGTGA

CGAGATGGTGGCCCGGTTCCAGTCCGGTGACGGCCCGCCATTTTCTGCTGTCGTTGAAGGC
GGGCGGTACCGGGCTGAACCTACCGCCGCCAATCATGTTGTGCACCTGGACCGCTGGTGGAA
CCCGGCGGTGAGAACAGGCGACGGACCGGGCGTTTCGGATCGGGCAGCGGCGCACGGTG
CAGGTCCGCAAGTTCATCTGCACCGGCACCCTCGAGGAGAAGATCGACGAAATGATCGAGGAG
5 AAAAAGGCGCTGGCCGACTTGGTGGTCACCGACGGCGAAGGCTGGCTGACCGAACTGTCCACC
CGCGATCTGCGCGAGGTGTTGCGGCTGTCCGAAGGCGCCGTCGGTGAGTAG

>Rv2110c prcB proteasome [beta]-type subunit 2 TB.seq 2369727:2370599 MW:30274

>emb|AL123456|MTBH37RV:c2370599-2369724, prcB SEQ ID NO:69

10 GTGACCTGGCCGTTGCCCGATCGCCTGTCCATTACTCTCTGGAACACCCGCTGTAGACC
TATCTTCTTTCACTGACTTCCTGCGCCGCCAGGCGCCGGAGTTGCTGCCGGCAAGCATCAGCG
GCGGTGCGCCACTCGCAGGCGGCGATGCGCAACTGCCGCACGGCACCACTTGTGCGGCTG
AAATACCCCGGCGGTGTTGTCATGGCGGGTGACCGGCGTTCGACGCAGGGCAACATGATTTCT
GGGCGTGATGTGCGCAAGGTGTATATACCGATGACTACACCGCTACCGGCATCGCTGGCACG
15 GCTGCGGTGCGGTTGAGTTTGCCCGGCTGTATGCCGTGGAAGTTGAGCACTACGAGAAGCTC
GAGGGTGTGCCGCTGACGTTTGCCGGCAAATCAACCGGCTGGCGATTATGGTGCGTGGAAT
CTGGCGGCGCGATGCAGGGTCTGCTGGCGTTGCCGTTGCTGGCGGGCTACGACATTCATGCG
TCTGACCCGCAGAGCGCGGGTTCGTATCGTTTCGTTGACGCCGCCGGCGGTTGGAACATCGAG
GAAGAGGGCTATCAGGCGGTGGGCTCGGGTTCGCTGTTGCGGAAGTCGTCGATGAAGAAGTTG
20 TATTCGCAGGTTACCGACGGTGATTCGGGGCTGCGGGTGCGGGTTCGAGGCGCTCTACGACGCC
GCCGACGACGACTCCGCCACCGGCGGTCCGGACCTGGTGCGGGGCATCTTCCGACGGCGGT
GATCATCGACGCCGACGGGGCGGTTGACGTGCCGGAGAGCCGGATTGCCGAATTGGCCCGCG
CGATCATCGAAAGCCGTTCCGGGTGCGGATACTTTCGGCTCCGATGGCGGTGAGAAGTGA

25 >Rv2118c - = B2126_C1_165 (83.6%) TB.seq 2377471:2378310 MW:30091

>emb|AL123456|MTBH37RV:c2378310-2377468, Rv2118c SEQ ID NO:70

GTGTCAGCAACCGGCCCATTCAGCATCGGCGAACGTGTTGAGCTACCGACGCTAAGGGGCGC
CGCTACACCATGTCGCTGACTCCCGGTGCCGAATTCCACACTCATCGTGGCTCGATCGCCACG
ACGCGGTGATCGGGTTGGAGCAAGGCAGCGTGGTCAAATCCAGCAACGGCGCCCTGTTCTGG
30 TGCTGCGCCCGCTGCTGGTGCAGTACGTGATGCGCGCGGGCCCGCAGGTGATCTATC
CCAAAGATGCGGCCAGATCGTGATGAGGGCGACATATTTCCCGGCGCGCGGGTGCTGGAG
GCAGGAGCCGGATCCGGTGCTCTGACCTTGCTTTGCTGCGGGCGGTTGGGCCGGCCGGACA
GGTGATCTCCTACGAACAGCGCGCCGATCATGCCGAACACGCCCGGCGCAATGTGAGCGGCTG
CTACGGCCAGCCGCCGGACAAGTGGCGACTGGTCGTCAGCGACCTCGCCGACTCCGAAGTGC
35 CCGACGGATCCGTTGATCGGGCCGTGCTCGACATGCTGGCGCCGTGGGAGGTGCTCGACGCG
GTATCGCGGCTGCTGGTGCCTGGCGGAGTGCTGATGGTCTACGTGGCCACCGTCACTCAGCTG
TCGAGGATCGTGGAGGCACTGCGGGCCAAGCAGTGCTGGACCGAACCGAGAGCCTGGGAGAC

GCTGCAGCGGGGCTGGAACGTCGTAGGGTTGGCGGTTCCGGCCGAGCATTGATGCGCGGGC
ATACCGCGTTCTGGTAGCAACGCGCCGGTTGGCGCCGGGGGCTGTGGCTCCGGCGCCGCTA
GGTCGTAAGCGCGAGGGACGCGACGGGTAG

5 >Rv2144c - TB.seq 2404166:2404519 MW:12028

>emb|AL123456|MTBH37RV:c2404519-2404163, Rv2144c SEQ ID NO:71

ATGCTGATCATTGCGCTGGTCTTGGCCCTGATTGGGCTCCTGGCCTTGGTGTTCGCGGTGGTCA
CCAGCAACCAGCTAGTGGCCTGGGTATGCATCGGGGCCAGCGTGCTGGGTGTGGCGTTGCTGA
TCGTGATGCGTTGCGAGAACGCCAGCAAGGTGGCGCGGACGAAGCTGATGGGGCTGGGGAA
10 ACGGGTGTGCGGAGGAAGCCGACGTCGACTACCCGGAGGAAGCCCCGAGGAGAGCCAAGC
CGTCGACGCCGGTGTATCGGCAGTGAGGAGCCATCGGAGGAGGCCAGCGAAGCGACCGAGG
AGTCGGCGGTATCGGCGGACCGAAGCGACGACAGCGCCAAGTAG

>Rv2146c - TB.seq 2405667:2405954 MW:10805

15 >emb|AL123456|MTBH37RV:c2405954-2405664, Rv2146c SEQ ID NO:72

TTGGTGGTGTTCCTCAGATCCTTGGGTTCCGCGTGTTTCATCTTCTGGCTGCTGCTGATCGCTCG
GGTCGTCGTTGAGTTCATCCGCTCGTTCAGCCGTGACTGGCGTCCCACCGGTGTCACCGTGGT
GATCTTGGAGATCATCATGTGATCACTGATCCGCCGGTGAAGGTGCTGCGCCGGCTGATCCC
GCAACTCACGATCGGCGCGGTCCGGTTCGACCTGTGATCATGGTGCTGCTGCTGGTTGCGTT
20 CATCGGTATGCAACTGGCGTTTGGTGCTGCGGCCTGA

>Rv2147c - TB.seq 2406119:2406841 MW:27630

>emb|AL123456|MTBH37RV:c2406841-2406116, Rv2147c SEQ ID NO:73

GTGAATAGTCACTGTAGTCACACCTTCATCACAGACAACAGATCTCCAGGGCTAGAAGGGGTC
25 ACGCAATGAGCACACTGCACAAGGTCAAGGCCTACTTCGGTATGGCTCCCATGGAGGATTACGA
CGACGAGTACTACGACGACCGCGCTCCCTCGCGCGGGTATGCGCGGCCCCGATTGACGACG
ACTACGGCCGCTACGATGGGCGCGACTACGACGACGCGCGCAGCGATTACGCGGTGACCTG
CGCGGTGAGCCGGCCGACTATCCACCACCGGGATATCGCGGCGGGTACGCGGACGAACCACG
TTCCGGCCCCGGGAGTTCGACCGCGCGGAGATGACACGGCCGCGCTTCGGATCGTGGCTGC
30 GCAACTCCACCCGCGGCGCGCTAGCGATGGACCCCCGCGGATGGCGATGATGTTTCGAGGAT
GGCCATCCGCTCTCGAAGATCACCACGCTGCGGCCCAAGGACTACAGCGAGGCTCGCACCATC
GGTGAGCGGTTCCGCGACGGCAGCCCGGTATCATGGATCTGGTGTCGATGGACAACGCCGAT
GCCAAGCGGCTGGTCGATTCGCGGCCCGCCTGGCCTTCGCGCTGCGCGGCTCGTTCGACAA
GGTCGCGACCAAGGTGTTCTGCTCTCGCCTGCAGACGTCGATGTGTCCCCGAGGAGCGCCG
35 CAGGATCGCCGAAACCGGGTTCTACGCCTACCAATAG

>Rv2148c - TB.seq 2406841:2407614 MW:27694

>emb|AL123456|MTBH37RV:c2407614-2406838, Rv2148c SEQ ID NO:74

ATGGCGGCGGATCTTTCGGCGTATCCAGACCGCGAATCGGAATTGACGCATGCGTTGGCGGCA
ATGCGATCGCGACTTGCGGCGGCCGCGGAGGCGGCGGGTCGCAATGTCGGCGAAATTGAACT
TCTACCGATTACCAAATCTTTCAGCAACCGATGTTGCGATTTTGTTCGATTGGGTTGTCGGTC
5 CGTTGGCGAATCGCGCGAACAGGAAGCTTCAGCCAAGATGGCCGAACCTTAATCGGTTGTTGGC
GGCTGCCGAGTTGGGTCACTCGGGGGGTGTGCACTGGCACATGGTGGGCCGATTCAACGCA
ACAAAGCCGGGTGCGTGGCTGCTGGGCGCACACCGCTCACTCGGTGGACAGCTCGCGGTTG
GTGACCGCGCTGGATCGGGCGGTTGTTGCGGCGCTGGCCGAACACCGTCGTGGCGAGCGGCT
GCGGGTTTACGTCCAGGTCAGCCTCGACGGTGACGGATCCCGGGGCGGCGTCGACAGCACGA
10 CGCCCCGCGCCGTAGACCGGATTTGCGCGCAGGTGCAGGAGTCAGAGGGCCTCGAACTGGTC
GGGTTGATGGGCATTCCGCCGCTGGATTGGGACCCGGACGAGGCCTTTGACCGGCTGCAATCG
GAGCACAACCGGGTGCGTGCGATGTTCCCGCACGCGATCGGTCTGTGCGCGGGCATGTCCAAC
GACCTTGAAGTCGCCGTCAAACATGGTTCGACCTGTGTGCGTGTGCGGTACCGCGCTATTGGGTC
CGCGGCGGTTACGGTCACCGTGA

15

>Rv2150c ftsZ TB.seq 2408386:2409522 MW:38757

>emb|AL123456|MTBH37RV:c2409522-2408383, ftsZ SEQ ID NO:75

ATGACCCCCCGCACAACTACCTGGCCGTCATCAAGGTCGTGGGTATCGGTGGTGGCGGTGTC
AACGCCGTCAACCGAATGATCGAGCAGGGCCTCAAAGGCGTGGAATTCATCGCGATCAACACC
20 GACGCCCAGGCGTTGTTGATGAGCGATGCCGACGTCAAACCTCGACGTGGGCCGCGACTCCACC
CGCGGGCTGGGCGCCGGCGCCGATCCGGAGGTGCGCCGTAAGGCCGCCGAGGACGCCAAGG
ACGAGATCGAAGAGCTGCTGCGCGGTGCCGACATGGTGTGTTGTCACCGCCGGCGAGGGGGGC
GGAACCGGCACCGGGGGGGGCACCCGTCGTGCGCAGCATCGCCGCAAGCTGGGCGCGTTGAC
CGTCGGTGTGGTCACCGGCCGTTCTCGTTGAGGGCAAGCGACGCAGCAATCAGGCCGAAAA
25 TGGCATCGCGGCGCTGCGGGAGAGTTGCGACACCCTCATCGTGATTCCCAACGACCGGTTGCT
GCAGATGGGAGATGCCGCGGTATCGCTGATGGATGCTTCCGTAGCGCCGACGAGGTGCTGCT
CAACGGCGTGCGAGGGCATCACCGACCTGATTACCACCCCGGGTCTAATCAACGTCGACTTCGC
CGACGTCAAGGGCATCATGTCCGGTGCCGGCACCGCACTGATGGGCATCGGCTCGGCCCGGG
CGGAAGGCCGGTCGCTCAAAGCGGCCGAGATCGCCATCAACTCGCCGTTGCTGGAAGCCTCGA
30 TGGAGGGCGCGCAAGGCGTGCTGATGTGATCGCCGGCGGCGACGCACTTGGGCTTGTTGAG
ATCAACGAGGCGGCCTCGTTGGTACAAGACGCCGCTCACCCCGATGCCAACATCATCTCGGC
ACCGTCATCGACGATTCGCTCGGTGACGAGGTGCGGGTGACCGTGATCGCGGCCGGCTTCGAC
GTCAGCGGTCCCGGCCGCAAGCCGGTGATGGGCGAGACCGGCGGCGCCACCGGATCGAGT
CAGCCAAGGCAGGCAAGCTCACCTCGACCTTGTTGAGCCGGTCGACGCCGTGAGCGTGCCGT
35 TGCACACCAACGGCGCAACCCTGAGCATCGGCGGTGATGACGACGATGTGACGTGCCGCCCT
TCATGCGCCGCTGA

>Rv2152c murC TB.seq 2410639:2412120 MW:51146

>emb|AL123456|MTBH37RV:c2412120-2410636; murC SEQ ID NO:76

GTGAGCACCGAGCAGTTGCCGCCGATCTGCGGCGGGTGACATGGTCCGGCATCGGCGGAGC
TGGCATGTCCGGGCATCGCCCGAATCCTGCTGGACCGCGGCGGGCTGGTCTCCGGGTCAGACG
5 CCAAGGAGTCCGCGGGTGTGCATGCGCTGCGGGCGCGGGGCGCGTTGATCCGGATCGGACAC
GACGCGTCTGCTGCTGGACCTGTTGCCCGGTGGCGCCACGGCGGTCTCACTACCCATGCCGC
CATCCCCAAAACCAACCCCGAGCTCGTGAAGCGAGGCGCCGCGGCATTCCCGTGGTGTGCG
GCCGGCCGTGCTGGCCAAGTTGATGGCCGGGCGCACACATTGATGGTCACCGGCACGCACG
GCAAGACAACGACGACGTCCATGCTGATCGTCCCTGCAGCACTGCGGGCTTGACCCGTCCT
10 TTGCGGTCCGCGGTGAGCTGGGGGAGGCCGGTACCAACGCCCATCACGGCAGTGGCGACTGT
TTCGTCGCCGAAGCCGACGAAAGCGATGGCTCGCTGTTGCAGTACACACCCACGTCGCGGTG
ATCACCAACATCGAGTCCGATCACCTGGACTTCTACGGCAGCGTCGAGGCGTATGTTGCGGTGT
TCGACTCCTTCGTGGAGCGCATTGTCCCCGGGGGTGCGCTGGTGGTGTGCACTGACGACCCCG
GAGGGGCCGCGCTGGCTCAGCGCGCGACTGAGCTGGGAATTCGAGTGTGCGATACGGGTGCG
15 GTGCCGGGTGAGACCATGGCAGCCACGTTGGTCTCGTGGCAGCAACAGGGGGTCGGCGCGGT
CGCACATATCCGGTTGGCCTCAGAACTAGCCACAGCACAGGGTCCCCGCGTGATGCGGCTGTC
GGTGGCCGGGCGACACATGGCGCTCAACGCGCTGGGAGCGCTGCTGGCCGCGGTGCAGATCG
GCGCCCCGGCCGACGAGGTGCTCGACGGGTGCGCGGCTTCGAAGGAGTCCGGCGACGATTC
GAACTGGTTGGGACCTGCGGCGTCGGAAAGGCGTCGGTGGCGGTGTTTCGATGACTACGCCAC
20 CACCCGACGGAGATCAGCGCGACACTGGCGGCGGCGCGCATGGTGTGCAACAGGGGCGACGG
TGGCCGCTGCATGGTTGTGTTTCAACCCCATTTGTATTGCGGACAAAGGCATTGCTGCTGAG
TTTGGGCGTGCGCTGAATGCCGCTGACGAGGTGTTTCGTA CTGACGTCTACGGAGCTCGTGAA
CAACCGCTGGCCGGTGTGAGCGGAGCCAGCGTCGCTGAGCACGTCACTGTGCCGATGCGCTA
CGTCCCGGATTTTTCGGCGGTGCGACAGCAAGTGGCCGCCGCCGCTAGTCCGGGCGACGTCAT
25 CGTCACGATGGGTGCCGGAGACGTGACCTTGCTGGGCCCGGAAATCCTGACCGCCCTTCGGGT
CCGGGCCAACC GAAGCGCCCCCGGCCGTCCGGGGGTGCTGGGATGA

>Rv2153c murG TB.seq 2412120:2413349 MW:41829

>emb|AL123456|MTBH37RV:c2413349-2412117; murG SEQ ID NO:77

GTGAAGGACACGGTCAGCCAGCCGGCCGGCGGGCGCGGGGCAACGGCGCCCCGGCCCGCG
30 ATGCCGCCTCGCCGCTTGTGGTTCTCGCCGTCTGCTGATTCCGTGTCCGTCTCTCGCCGG
CGGCGGGACCGCCGGGCACGTGAGCCCGCCATGGCCGTGCGCGACGCCTTGGTCTCGCTTGG
ATCCGCGCGTCCGGATTACCGCGTTGGGCACCTCCGTGGACTAGAGACCAGGCTGGTGCCCC
AGCGCGGCTACCACCTGGAGCTGATCACGGCGGTGCCGATGCCGCGCAAGCCCGGCGGGCGAC
35 CTGGCCCGGCTGCCGTGCGGGGTGTGGCGCGCCGTCCGGGAGGCCCGGGACGTGCTCGACG
ATGTCGACGCGCGACGTGCTGCTCGGTTTCGGTGGGTACGTGCGGCTACCGGCTTACCTAGCCG
CTCGCGGCCTGCCTTTGCCGCCCGGCGCCGGCGCCGGATCCCGGTGGTGTATCCACGAAGCC

AACGCCAGGGCGGGACTGGCCAACCGGGTCGGCGCCCATACCGCGGACCGGGTGCTCTCCGC
GGTGCCCGATTCCGGGCTGCGGCGCGCCGAGGTGGTTGGGGTCCCGGTCCGTGCGTCGATCG
CCGCGCTGGACCGCGCGGTGCTGCGAGCCGAGGCGCGGGCACACTTCGGCTTCCCCGACGAC
GCGCGGGTGCTGCTGGTGTTTCGGGGGTTTCGAGGGCGCGGTCTCGCTCAACCGGGCGGTGTC
5 CGGCGCCCGCCGCCGACCTGGCCGCCGCCGGTGTTTGCCTGCTGCATGCCATGGACCCAGA
ACGTGCTGGAGTTGCGCCGTCGGGCTCAAGGTGACCCACCGTACGTGGCGGTGCCCTATTTGG
ACCGGATGGAGCTGGCCTACGCCGCCGCCGATCTGGTGATCTGCCGGGCCGGGGCGATGACG
GTCGCCGAAGTATCCGCCGTCGGTCTGCCGCCATCTACGTGCCGCTGCCGATCGGCAACGGT
GAACAGCGGCTGAATGCGTTGCCGGTAGTCAATGCCGGCGGCGGCATGGTGGTCGCCGACGC
10 CGCCCTGACCCCCGAGTTGGTGGCCCGCCAGGTTGCCGGGCTGCTACCGACCCCCGCGCGGC
TGGCCGCGATGACCGCGGCCCGCAGCCAGGGTGGBACATCGCGATGCCGCGGGCCAGGTGGC
CCGGGCCGCGCTGGCCGTCGCCACCGGGGCCGGTGCCAGGACAACGACGTGA

>Rv2154c ftsW TB.seq 2413349:2414920 MW:56306

>emb|AL123456|MTBH37RV:c2414920-2413346, ftsW SEQ ID NO:78

15 GTGCTAACCCGGTTGCTGCGTCGGGGCACACGCGACACCGACGGTCCCAGACTCGAGGGGC
CGAGCCGGTCGAGGGGCAGCGGACGGGCCCGGAAGAAGCCTCTAACCCGGGTTCCGCGAGG
CCCCGCACCCGTTTCGGTGCTGGCTGGGCCGTCCGATGACCTCGTTTCACCTCATCATCGCC
GTTGCCGCATTGCTGACCACCCCTGGACTGATCATGGTGCTGTCGGCATCGGCGGTGCGGTCC
20 TACGACGACGACGGATCGGCTTGGGTGATCTTCGGCAAGCAGGTCTTGTTGGACGCTTGTGGGT
CTTATCGGCGGCTATGTCTGTCTGCGGATGTCGGTGCGGTTTCATGCGGCGCATCGCCTTCTCCG
GTTTCGCGATCACCATCGTGATGCTGGTGCTGGTGCTGGTGCCGGGGATCGGCAAGGAGGCCA
ACGGCTCGCGCGGCTGGTTCGTGGTCGCGGGCTTCTCGATGCAGCCCTCTGAGCTGGCTAAGA
TGGCGTTCGCCATCTGGGGAGCGCATCTGCTGGCCGCCCGCGCATGGAACGGGCTTCACTG
25 CGCGAGATGCTGATTCACCTGGTGCCGGCCGCCGTCGTTGCGCTGGCGCTGATCGTGCCCGAG
CCCGACCTCGGACAGACCGTGTCGATGGGCATCATCTTGTTGGGCCTGCTGTGGTATGCGGGG
CTGCCGCTGCGCGTCTTCCTCAGCTCACTGGCGGCGGTGCTGCTCTCGGCCGCCATCCTGGCG
GTGTCCGCGGGCTACCGATCCGACCGGGTGCGGTCGTGGCTCAACCCCGAAAACGATCCGCAA
GACTCCGGCTACCAGGCCCGACAGGCAAAGTTCGCGCTGGCTCAAGGTGGCATTTTCGGCGAC
30 GGTCTGGGCCAAGGCGTGCCCAAGTGAAGTACTTGCCCAACGCCACAACGACTTCATTTTCG
CCATCATCGGCGAAGAGCTGGGTCTCGTCGGCGCGCTCGGACTGCTGGGGCTATTCCGATTGT
TCGCCTACACCGGCATGCGCATCGCTAGCCGGTCCGCCGACCCGTTCTGCGGCTGCTGACCG
CCACCACGACACTGTGGGTGCTGGGACAGGCGTTCATCAACATCGGCTATGTGATCGGGCTGC
TGCCCGTCACCGGCCTGCAGCTGCCGCTCATCTCCGCCGGTGGAACCTCCACGGCCGCAACAC
35 TTTGCTGATAGGCATCATCGCCAACGCGGCTCGCCACGAACCGGAGGCGGTGGCCGCGCTG
CGGGCTGGGCGCGACGACAAGGTGAACCGGTTGCTGCGGCTGCCGCTGCCCGAGCCGTATCT
GCCCCCTCGTCTCGAGGCGTTTCGTGACCGCAAGCGCGCCAACCCGCAACCGGCCCAAACGCA

GCCCGCGCGGAAGACCCCCGCACGGCGCCCGGACAGCCTGCCCGGCAGATGGGCCTGCCC
CCGCGACCCGGCTCGCCCCGCACGGCCGATCCGCCGGTTCGTCGATCAGTGCATCATGGAGCT
GGCCAGCGGTACGCGGGCCAGCGTCGCACACGGCGCGTTCCGGGCATTGGAAGGTCAGCGTTA
CGGGTGA

5

>Rv2155c murD TB.seq 2414935:2416392 MW:49314

>emb|AL123456|MTBH37RV:c2416392-2414932, murD SEQ ID NO:79

GTGCTTGACCTCTGGGGCCGGGTGCGCCCGTGTGGTAGCCGGTGGCCGGGTGACCGGTCA
GGCGGTGGCCGCGGTGCTGACTCGGTTTGGTGCACGCCGACGGTGTGCGACGACGATCCGG
10 TCATGCTGCGACCCGACGCCGAACGTGGGCTGCCGACCGTTAGTTCCTCGGACGCGGTGCAGC
AGATAACCGGGTATGCGCTGGTGGTGCACAGTCCCGGCTTCTCGCCCGCAACCCCGCTACTGG
CCGCGGCCGCGGCGGCGGGGGTGCCGATCTGGGGTGACGTGGAGTTAGCCTGGCGGCTAGA
CGCAGCGGGCTGCTACGGACCGCCGCGCAGCTGGCTGGTGGTGACCGGCACCAACGGCAAGA
CCACCACGACGTCGATGCTGCACGCCATGCTGATCGCCGGTGGCCGCCGCGCCGTGCTGTGC
15 GGCAATATCGGCAGTGCGGTGCTGGATGTGCTGGACGAGCCGGCCGAGCTGCTGGCCGTGGA
GTTGTCCAGTTTCCAGCTGCACTGGGCGCCGTGCTGCGGCCCGAGGCCGGCGCGGTGCTCA
ACATTGCCGAAGACCACCTGGACTGGCATGCCACGATGGCCGAATACACCGCGGCCAAGGCCC
GGGTGCTGACCGGCGGGGTAGCGGTGGCCGGGCTGGATGACAGCCGAGCGGCCGCACTGCT
GGACGGCTCACCGGCGCAGGTGCGGGTCGGCTTCCGGCTCGGCGAGCCGGCCGCGCGGGAA
20 CTGGGCGTGCGCGACGCCACCTGGTGCATCGCGCCTTCTCCGACGACTTGACGCTGCTGCCG
GTGCGCTCGATACCGGTGCCAGGTCCGGTCGGCGTGCTTGACGCCCTGGCCGCGGCGGCGCT
GGCCCGCTCGGTGCGGGTGCCCGCCGGTGCATCGCCGACGCGGTACGTCGTTTCGAGTGG
GCCGACACCGCGCCGAGGTGGTGGCCGTTGCCGACGGCATCACCTACGTGGACGACTCCAAG
GCCACCAACCCGCACGCCGCGCGGGCTTCGGTGCTTGATACCCGAGGGTGGTATGGATCGC
25 CGGTGGCCTGCTCAAGGGCGCGTCGCTTACGCCGAGGTTGCGGCGATGGCGTCGCGGCTGG
TCGGTGCGGTGCTGATCGGCCGGGATCGCGCAGCGGTTGCCGAGGCGTTATCACGACACGCG
CCCGATGTCCAGTCGTTAGGTTGTGGCAGGCGAGGATACTGGTATGCCTGCGACTGTTGAG
GTTCTGTTGCTTGTGTTCTAGATGTGGCAAAGATGACAAAGCCGGTGAGACCGTTGGCGCTG
CCGTGATGACCGCTGCGGTGGCCGCGGCCCGGCGGATGGCCCAACCCGGTGACACCGTGCTG
30 CTGGCACCGGCCGCGCCTCATTCGACCAGTTCACCGGTTATGCCGACCGGGGCGAGGCATTC
GCGACCGCGGTCCGCGCGGTGATCCGGTAG

>Rv2156c murX TB.seq 2416397:2417473 MW:37714

>emb|AL123456|MTBH37RV:c2417473-2416394, murX SEQ ID NO:80

ATGAGGCAGATCCTTATCGCCGTTGCCGTAGCGGTGACGGTGTCCATCTTGCTGACCCCGGTG
CTGATCCGGTTGTTCACTAAGCAGGGCTTCGGCCACCAGATCCGTGAGGATGGCCCGCCAGC
35 CACCACACCAAGCGCGGTACGCCGTCGATGGGCGGGGTGGCGATTCTGGCCGGCATCTGGGC

GGGCTACCTGGGCGCCACCTAGCGGGCTGGCGTTTGACGGTGAAGGCATCGGCGCATCGG
GTCTGTTGGTGCTGGGCCTAGCCACCGCTTTGGGCGGCGTCGGGTTTCATCGACGATCTGATCA
AGATCCGCAGGTCGCGCAATCTCGGGTTGAACAAGACGGCCAAGACCGTCGGGCAGATCACCT
CCGCCGTGCTGTTTGGCGTGCTGGTGCTGCAGTTCGGGAATGCTGCCGGCCTGACACCGGGCA
5 GCGCGGATCTGTCCTACGTGCGTGAGATCGCCACCGTCACATTGGCGCCGGTGCTGTTCTGTG
TGTCTGCGTGGTCATCGTCAGCGCCTGGTCGAACGCGGTCAACTTCACCGATGGCCTGGACG
GGCTGGCCGCCCGGCACCATGGCGATGGTCACCGCCGCCTACGTGCTGATCACCTTCTGGCAGT
ACCGCAACGCGTGCGTGACGGCGCCGGGCTGGGCTGCTACAACGTGCGCGACCCGCTGGAC
CTGGCGCTCATCGCGGCCGCAACCGCTGGCGCCTGCATCGGTTTTTTGTGGTGGAACGCCGCG
10 CCCGCCAAGATCTTCATGGGTGACACTGGGTGCGTGCGCTTGGGCGGCGTCATCGCGGGGTTG
TCGGTGACCAGCCGCACCGAGATCCTTGCGGTGGTGCTGGGTGCGCTGTTCTGTCGCCGAGATC
ACCTCGGTGGTGTGCAAATCCTGACCTTCGGGACCACCGGGCGCCGGATGTTTCGGATGGCG
CCCTTCCACCACCATTTTCGAGTTGGTCGGTTGGGCTGAAACCACGGTCATCATCCGGTTCTGGC
TGCTACCGCGATCACCTGCGGTCTGGGCGTGGCCTTGTTCTACGGTGAGTGGCTTGCCGCGG
15 TCGGTGCCTGA

>Rv2157c murF TB.seq 2417473:2419002 MW:51634

>emb|AL123456|MTBH37RV:c2419002-2417470, murF SEQ ID NO:81

ATGATCGAGCTGACCGTCGCGCAGATCGCCGAGATCGTCGGGGGCGCAGTGGCCGATATCTCC
20 CCGCAAGACGCCGCGCACCGCCGCGTCACCGGGACCGTCGAGTTCGACTCGCGCGCCATCGG
CCCGGGCGGGGCTGTTCTCGCCCTGCCGGGGGCGCGCGCCGACGGCCACGACCATGCCGCG
TCGGCGGTAGCCGCGGGCGCCGCCGTCGTGCTGGCCGCCCGCCCGGTGGGGGTGCCGGCCA
TCGTGGTTCCGCCAGTGGCGCGCCGAACGTATTGGCCGGCGTCCTCGAGCACGACAACGAC
GGGTGCGGGGCGGCGGTGCTGGCCGCGCTGGCCAAGCTGGCCACCGCGGTGGCCGCGCAGT
25 TGGTGGCCGGCGGGCTCACCATCATCGGGATCACCGGCTCGTCGGGCAAGACGTCGACCAAG
GACCTGATGGCCGCCGTGCTGGCCCCGCTGGGGGAGGTGGTGGCCCCGCCCGGATCGTTCAA
CAACGAGCTGGGTACCCGCTGGACGGTGCTGCGCGCGACGCGGCGCACCGACTACCTGATTTT
GGAGATGGCGGCACGCCATCACGGCAACATCGCCGCGCTCGCCGAGATCGCGCCCCCGTCGA
TCGGAGTCGTGCTCAACGTGGCACCGCACATTTGGGTGAGTTCGGCTCCCGCGAGGTCATCG
30 CACAGACCAAAGCCGAAGTGGCGCAGGCTGTTCCGCATTCCGGAGCGGTGTCCTCAACGCTG
ATGACCCCGCGGTGGCGGCGATGGCCAAGCTGACCGCGGCCCGGGTGGTGCGGGTCAGCCG
GGACAACACCGGTGACGTTTGGGCGGGGCGCGGTGTCGCTGGACGAATTGGCCAGGCCGCGCT
TTACGCTGCATGCCACGATGCCAAGCCGAGGTCCGACTCGGGGTCTGCGGCGACCAACG
GTCACCTAACGCGCTGTGCGCCGCGGCGGTGCGCTGGAGTGTGGGGCCAGCGTTGAACAGGT
35 CGCGGCCGCGCTGACCGCGGCGCCGCCGGTGTGCGGCGCATCGGATGCAGGTGACCACCCGC
GGCGACGGGGTGACGGTGATCGACGACGCCTACAACGCCAACCCCGACTCCATGCGGGCCGG
GCTGCAGGCGCTGGCCTGGATCGCGCACCAACCCGAGGCCACCCGCCGAGCTGGGCGGTGC

TGGGTGAGATGGCCGAGCTGGGTGAGGACGCGATAGCCGAGCACGATCGCATCGGCCGGCTC
GCGGTGCGCTTAGATGTGTCTCGACTCGTTGTCGTGGGAACCGGGAGGTGATCAGCGCCATG
CACCACGGAGCGGTCTCTGGAGGGGGCGTGGGGCTCGGGGGAAGCCACTGCTGATCACGGTGC
GGATCGCACGGCCGTCAATGTGGCCGACGGTGACGCCGCCCTGGCACTACTGCGCGCCGAGC
5 TGCACCCCGGGGATGTGGTCTTGGTCAAGGCCTCGAACGCGGCCGGGCTGGGTGCGGTGGCC
GATGCATTGGTTCGACAGACGACATGCGGGAGTGTGCGCCCATGA

>Rv2158c murE TB.seq 2419002:2420606 MW:55310

>emb|AL123456|MTBH37RV:c2420606-2418999, murE SEQ ID NO:82

10 GTGTCATCGCTGGCCCCGAGGATCTCGCGGCGGCGAACGGAGGTGGCGACACAGGTGGAGGC
TGCGCCCACTGGCTTGCGCCCCAACGCCGTGCTGGGCGTTGCGTTGGCCGCACTGGCCGATCA
GGTCGGCGCGGCCCTGGCCGAGGGTCCAGCTCAGCGTGCCGTACCGAGGACCGGACGGTCA
CCGGGGTACGCTGCGCGCCCAGGACGTGTACCCGGTGACCTGTTGCGCGCCCTGACCGGC
TCGACCACCCACGGGGCCCCGCCACGTGCGCGACGCGATCGCACGCGGCGCCGTGCGGGTGCT
15 CACCGACCCCGCCGGGTGCGCGAGATCGCCGGACGAGCGGCCGTGCCCGTGTTGGTGCACC
CCGCACCCCGCGCGTGCTCGGCGGCTTGGCCGCCACCGTGTACGGGCATCCGTCCGAGCGG
TTGACGGTTATCGGGATCACCGGAACGTCCGGCAAGACCACCACCTATCTGGTCGAGGCC
GGGTTACGGGCTGCCGGACGCGTCGCCGGGCTGATCGGCACCATCGGCATCCGCGTCGGCGG
CGCCGACCTTCCAGCGCGCTGACCACCCCGGAGGCCCCACGCTGCAGGCGATGCTGGCGG
20 CGATGGTGAACGCGGGGTGGACACCGTGGTCATGGAGGTGTCCAGCCACGCGCTGGCGCTG
GGCCGGGTGGACGGCACCCGGTTCGCCGTGCGCGCCTTACCAATCTCTCCCGTGACCACCTG
GATTTCCACCCAGCATGGCCGACTACTTCGAGGCCAAGGCGTCATTGTTGATCCGGACTCGG
CACTGCGCGCCCCGCACCGCCGTGGTGTGCATCGACGACGACGCCGGGCGCGCGATGGCGGC
GCGGGCCGCGACGCGATCACCGTCAGCGCCGCCGACCGGCCCGCACACTGGCGCGCCACG
25 GATGTGGCGCCACGGACGCGGGCGGGCAACAATTCACCGCCATCGACCCCGCCGGCGTAGG
GCATCACATCGGAATCCGGCTACCGGGCCGCTACAACGTGCGCAATTGCCTGGTCGCCCTGGC
GATTCTGGACACCGTCGGGGTCTCCCCGGAACAGGCGGTGCCGGGCCTGCGTGAGATCCGGG
TCCCGGGGCGGCTCGAGCAGATCGACCGCGGCCAGGGCTTTCTCGCGCTGGTCGACTACGCG
CACAAACCGGAAGCGCTGCGGTGCGTGCTGACCACCTTGGCGCACCCGGACCGCCGGCTGGC
30 GGTGGTGTTGCGCGCCGGCGCGGATCGTGACCCGGGCAAGCGGGCCCCGATGGGCGCGGATA
GCCGCGCAGCTGGCCGACTTGGTGGTCGTACCGACGACAACCCCGGTGACGAAGATCCAC
GGCGATCCGCCGCGAAATCCTGGCTGGGGCGGCCGAAGTCGGCGGTGATGCCAGGTGCTCG
AGATCGCAGACCGGCGGGACGCGATCCGGCACGCGGTTGCCTGGGCGCGCCCCGGCGACGT
GGTGCTCATCGCCGGCAAAGGCCACGAGACCGGGCAACGCGCGCGGGCGGGTCCGCCCCG
35 TTCGACGACCGGGTGGAGCTGGCTGCCGCGCTAGAGGCCCTCGAGCGGCGCGCATGA

>Rv2159c - TB.seq 2420632:2421663 MW:36377

>emb|AL123456|MTBH37RV:c2421663-2420629, Rv2159c SEQ ID NO:83

ATGAAATTTGTCAACCATATTGAGCCCGTCGCGCCCCGCCGAGCCGGCGGGCGGGTCGCCGAG
GTCTATGCCGAGGCCCGCCGCGAGTTGGGCCGGCTGCCGAGCCGCTCGCCATGCTGTCCCC
GGACGAGGGACTGCTCACCGCCGGCTGGGCGACGTTGCGCGAGACACTGCTGGTGGGCCAGG
5 TCCGCGTGGCCGCAAGGAAGCCGTCGCCGCCGCGTCGCGGCCAGCCTGCGCTGCCOCTGG
TGCCTCGACGCACACACCACCATGCTGTACGCGGCAGGCCAAACCGACACCGCCGCGGGAT
CTTGCCCGGCACAGCACCTGCCGCCGGTGACCCGAACGCGCCGTATGTGGCGTGGGCGGCAG
GAACCGGGACACCGCGGGACCGCCGGCACC GTTCGGCCCCGGATGTCGCCGCCGAATACCTG
GGCACCGCGGTGCAATTCACCTTCATCGCACGCCTGGTCCTGGTGCTGCTGGACGAAACCTTC
10 CTGCCGGGGGGCCCCGCGCGCCCAACAGCTCATGCGCCGCGCCGGTGGACTGGTGTTCGCCCCG
CAAGGTGCGCGCGGAGCATCGGCCGGGCGCTCCACCCGCCGGCTCGAGCCGCGAACGCTG
CCCGACGATCTGGCATGGGCAACACCGTCCGAGCCCATAGCAACCGCGTTCGCCGCGCTCAGC
CACCACCTGGACACCGCGCCGCACCTGCCGCCACCGACTCGTCAGGTGGTCAGGCGGGTCGT
GGGGTCGTGGCACGGCGAGCCAATGCCGATGAGCAGTCGCTGGACGAACGAGCACACCGCCG
15 AGCTGCCCCGCGACCTGCACGCGCCCAACCGTCTTGCCCTGCTGACCGGCCTGGCCCCGCAT
CAGGTGACCGACGACGACGTCGCCGCGGCCCGATCCCTGCTCGACACCGATGCGGCGCTGGT
TGCGCCCTGGCCTGGGCGGCCTTCACCGCCGCGCGGCGCATCGGCACCTGGATCGGCGCCG
CCGCCGAGGGCCAGGTGTGCGGGCAAAACCCGACTGGGTGA

20 >Rv2163c pbpB TB.seq 2425049:2427085 MW:72506

>emb|AL123456|MTBH37RV:c2427085-2425046, pbpB SEQ ID NO:84

GTGAGCCGCGCCGCCCCAGGCGGGCCAGTCAGTCGACGCGACCGGCGCGCGGTTT
GCGCCGGCCACCGGGAGCCCAGGAGGTTGGGCAACGCAAACGGCCCGGCAAAACGCAGAAAG
CCCGGCAAGCCCAGGAAGCCACGAAATCCCGCCCTGCGACACGGTCAGACGTCGCACCCGCG
25 GGTGCTCGACTCGTGCGAGGCGCACCCGGCAGGTGGTGGACGTCGGGACGCGCGGTGCGTC
GTTCTGCTTTTCGGCATCGGACCGGAAACGCGGTTCATCTTGGTGTTGATGTTGGTCGCGGCAACA
CAATTGTTCTTTCTGCAGGTATCATATGCCGCGGGCCTGCGTGCGCAGGCGGGCCGCAACTC
AAGGTCACCGACGTCCAGCCAGCGGCTCGCGGCAGCATCGTCGACCGCAACAATGACCGGCTC
GCGTTCACCATCGAGGCGCGTGCCCTGACGTTCCAGCCGAAGCGGATTCGGCGGCAATTGGAA
30 GAGGCCAGGAAGAAGACGTCGGCTGCACCCGACCCGAGCAGCGCCTGCGCGATATCGCCCA
GGAGGTCGCCGGCAAGCTGAACAACAAGCCAGATGCCGCGGCCGTGCTGAAGAAGCTGCAAA
GCGACGAGACCTTCGTCTACTTGCGCGGTGCGGTGCGACCCGGCTGTGCGCAGCGCGATCTGCG
CGAAGTATCCCGAGGTGGTGCGGAAAGACAGGATCTGCGTCAGTACCCGGGTGGGTGCGCTG
GCGGCAACGTCGTGCGGTGGCATCGACTGGGATGGTCATGGGCTGCTGGGTCTGGAGGACTCC
35 CTGGATGCGGTGCTGGCCGGAACCGACGGATCGGTACCTACGACCGTGGGTGACACGGCGT
CGTCATCCCCGGCAGCTACCGGAATCGGCACAAGGCGGTCCACGGTTCACCGTCTGTCTCAC
CCTCGACAACGACATCCAAGTTCTACGTGCAGCAGCAGGTGCAGCAGGCCAAGAACCTATCGGG

GGCTCACAACGTCTCGGCCGTCTGCTCCTGGACGCCAAGACCGGCGAGGTGCTCGCGATGGCCA
ACGACAACACCTTCGACCCGTGCAAGACATCGGGCGCCAGGGCGACAAGCAGTTGGGCAACC
CGGCGGTGTCTGTCGCCCTTCGAGCCGGGCTCGGTGAACAAGATCGTCGCCCGTCCGCGGTC
ATCGAGCACGGGTTGAGCAGCCCCGACGAGGTGCTACAGGTGCCTGGCTCGATCCAGATGGG
5 CGGTGTTACCGTGCATGACGCTTGGGAGCACGGCGTGATGCCCTATACCACCACGGGGGTGTT
CGGAAAGTCCTCCAACGTCCGCACGCTGATGCTTTCCCAACGTGTCCGACCGGAACGCTATTAC
GATATGCTCCGCAAGTTCGGGTTGGGACAGCGCACCGGCGTGGGCCTGCCCGGTGAGAGCGC
CGGACTGGTGCCGCCAATCGACCAGTGGTCGGGCAGTACGTTTCGCTAATCTTCCTATTGGCCAA
GGTCTTTGATGACTTTGCTGCAGATGACCGGCATGTACCAGGCCATCGCCAACGATGGAGTGC
10 GGGTACCCCCACGCATTATCAAGGCCACCGTCGCACCCGACGGCAGCCGAACCGAAGAACCGC
GCCCCGACGACATTGCGGTGGTGTCCGGCGCAGACCGCCAGACCGTGCGCCAGATGCTGCGT
GCCGTGGTGCAACGCGATCCGATGGGCTACCAGCAGGGTACCGGGCCGACGGCCGGGGTGCC
CGGCTATCAGATGGCCGGCAAGACCGGTACCGCGCAGCAGATCAACCCTGGCTGCGGGTGCTA
CTTCGACGACGTGTATTGGATCACCTTCGCCGGAATCGCCACTGCCGACAATCCCCGCTACGTG
15 ATCGGCATCATGTTGGACAACCCGGCGCGCAACTCCGACGGCGCGCCTGGGCACTCGGCCGC
CCCGCTGTTCCACAACATCGCGGGCTGGCTGATGCAGCGCGAAAACGTCCCGCTGTCACCCGA
TCCCGGGCCTCCTTTGGTCTTGACAGGCCACCTAG

>Rv2165c - TB.seq 2428236:2429423 MW:42498

20 >emb|AL123456|MTBH37RV:c2429423-2428233, Rv2165c SEQ ID NO:85
GTGCAAACCCGTGCACCGTGCTCTGCCCCGAAGCGACCCCTGGCGTACTTCCCCAACGCCAGG
TTCGTGTCTTCGGACAGGGACCTCGGTGCAGGGGCGGCGCCTGGAATAGCCGCGTCCCGAAGT
ACGGCTTGCCAGACCTGGGGAGGTATCACGGTGGCTGATCCAGGTTCCGGGGCCAACCGGTTTC
GGTCATGTGCCGGTATTGGCGCAACGTTGCTTCGAACTGCTTACCCCCGCACTAACCCGCTACT
25 ATCCAGACGGCTCGCAGGCGGTCTTCTCGACCGGACCATCGGCGCGGGCGGGCATGCGGAG
CGGTTTTTGGAGGGATTGCCGGGTCTGCGCCTGATCGGGCTCGACCGTGACCCAACCGCTCTG
GACGTGCGCGGGTCTCGGCTGGTGCGATTGCTGACCGACTTACCCTGGTGACACCCGCTAT
GACTGTCTGGGCGCAGCGCTGGCTGAATCCGTTATGCCGAGTGGGATCAGTCGACGGAATC
CTGTTGATCTCGGCGTCTCATCCATGCAGCTCGACCGCGCCGAGCGGGGCTTCGCCTACGCC
30 ACGGACGCGCCATTGGACATGCGGATGGACCCGACGACGCCGTTGACCGCAGCTGACATTGTC
AACACTTACGACGAGGCGGCACTAGCCGACATCCTGCGTCGCTACGGAGAGGAGCGGTTTGCT
CGGCGCATCGCTGCCGGTATCGTCCGCCGACGCGCAAAAACCCCGTTACCTCGACCGCCGAA
CTGGTTGCCCTGCTGTACCAGGCGATTCCAGCTCCGGCCCCGGCGTGTCCGGCGGCATCCAGCC
AAGCGAACATTCCAGGCGCTGCGCATGCGGGTCAACGATGAGCTGGAATCGCTGCGCACGGCC
35 GTTCCTGCCGCGCTGGATGCCCTCGCTATCGGTGGGCGCATCGCGGTGCTGGCCTACAGTCG
CTAGAGGACAGGATCGTCAAACGGGTGTTCCGCCGAGGCAGTCGCGTGGGCCACCCCTGCGGG
ACTTCGGTTCGAACTTCCCGGCCATGAGCCGCGATTCCGTTGTTAACGCACGGCGCCGAACG

AGCGAGTGTGGCTGAGATCGAACGCAATCCCCGCAGTACTCCAGTGCGGTTGCGGGCCCTGCA
ACGAGTCGAGCACCGGGCGCAATCGCAGCAATGGGCAACCGAGAAGGGTGATTCATGA

>Rv2166c - TB.seq 2429428:2429856 MW:15912

5 >emb|AL123456|MTBH37RV:c2429856-2429425, Rv2166c SEQ ID NO:86

ATGTTTCTCGGCACCTACACGCCAACTCGACGACAAGGGGCGGCTGACGCTGCCGGCCAAG
TTTCGCGACGCGTTGGCAGGGGGGTTGATGGTCACCAAGAGCCAAGATCACAGCCTGGCCGTT
TACCCGCGGGCGGCGTTTCGAGCAGCTGGCGCGCCGGGCCAGCAAGGCGCCACGAAGCAACC
CCGAGGCGAGAGCGTTCCTACGTAATCTCGCCGCCGGTACCGACGAACAGCATCCCGACAGTC
10 AAGGCCGGATCACCTTGTGCGCCGACCACCGCCGCTACGCAAGCCTTTCCAAGGACTGTGTGG
TGATCGGCGCGGTGACTATCTCGAGATCTGGGATGCGCAAGCCTGGCAGAACTACCAACAAAT
CCATGAAGAGAACTTCTCCGCGGCCAGCGATGAAGCACTCGGTGACATCTTCTGA

>Rv2197c - TB.seq 2461505:2462146 MW:22481

15 >emb|AL123456|MTBH37RV:c2462146-2461502, Rv2197c SEQ ID NO:87

ATGGTGAGCAGATATTCGCATACCGGCGTGGGCCGGATGTAATCTCGCCGGACGTCATCGAT
CGCATCCTGGTTGGGGCATGTGCCGCGGTGTGGCTGGTGTTACCGGCGTGTGCGGTGGCCGC
CGCTGTGCCCTGATGGACCTGGGTAGGGGCTTCCACGAGATGGCCGAAACCCGCACACCAC
GTGGGTGCTGTACGCCGTAATTGTGGTCTCCGCACTGGTCATCGTGGGCGCGATACCGGTGCT
20 GTTGCGAGCTCGCCGCATGGCTGAGGCCGAGCCGCGACGAGGCCGACGGGTGCATCCGTGC
GGGGCGGGCGATCGATCGGATCCGGGCATCCGGCGAAACGCGCTGTGGCCGAGTCGGCACCC
GTACAGCACGCGGATGCATTCGAGGTGGCCGCCGAGTGGTCCAGTGAGGCGGTGGACCGGAT
CTGGTTGCGCGGGACAGTCGTGTTGACCAGTGCGATTGGCATTGCGTTGATTGCCGTGGCGGC
GGCGACCTACCTCATGGCGGTGCGTCACGACGGGCCATCTTGGATCAGCTACGGGTTGGCCGG
25 GGTGGTCACCGCGGGCATGCCGGTGATCGAGTGGCTATACGCTCGGCAGCTGCGCCGGGTGG
TGGCGCCCCAGTCCAGTTAG

>Rv2198c - TB.seq 2462149:2463045 MW:30955

>emb|AL123456|MTBH37RV:c2463045-2462146, mmpS3 SEQ ID NO:88

30 ATGAGCGGGCCGAATCCCCGGGACGGGAACCTGACGAACCCGAATCGGAACCCGTCAGCGA
CACGGGCGACGAACGGGCTTCCGGCAACCACTTGCCGCCCGTCGCCGGGGGCGGCGACAAAC
TGCCCAGTGACCAGACGGGCGAGACCGACGCATATTCTCGGGCATACTCTGCCCCGGAATCCG
AGCACGTACCGGCGGGCCCGTATGTGCCAGCCGATCTCAGGCTCTATGACTACGACGACTATG
AGGAGTCGTCCGACCTGGACGACGAACTGGCCGCTCCGCGCTGGCCGTGGGTGGTCGGTGTC
35 GCCGCCATAATTGCCGCCGTTGCGCTCGTGGTTTCGGTGTGCTTGCTCGTCACGCGACCACATA
CCAGCAAACCTGCCACCGGCGACACTACGTCCTCTGCACCGCCCGTGACGAGACGAAATCACGA
CCACCAAGCCGGCGCCGCCACCGCCGCCACCAGCCCCACCGCCACCACCGAGATCCCGACA

GCGACGGAGACACAGACGGTCACTGTGACGCCGCCACCACCGCCCCACCGGCGACAACCAC
GGCGCCGCCGCCGGCGGACCACCAACGGCGGGCGGCACCGCCGCCACGACCACCACGCCG
ACCGGTCCGCGGCAAGTCACCTATTCGGTGACCGGTACCAAGGCGCCGGGTGACATTATCTCG
GTGACTTACGTGATGCCGCCGGGCGCCGACGGACACAGCACAATGTGTACATCCCGTGCTCC
5 ATGACGGTCACCCCGATCTCGCAATCCGACGTTGGCTCGGTGGAGGCCTCCAGCCTTTTCCGG
GTCAGCAAACCTCAACTGCTCGATCACCACGAGCGACGGAACGGTGCTCTCATCGAACTCCAACG
ATGGACCGCAAACGAGCTGCTGA

>Rv2199c - TB.seq 2463234:2463650 MW:14866

>emb|AL123456|MTBH37RV:c2463650-2463231, Rv2199c SEQ ID NO:89

10 ATGCATATCGAAGCCCCGACTGTTTGAGTTTGTGCGCCGCGTCTTCGTGGTGACGGCGGTGCTGT
ACGGCGTGTTGACCTCGATGTTGCCACCGGTGGTGTGAGTGGGCTGGCACCCTGCGCTGG
CGCTTACCGGCGGCATGGCGTTGATCGTCGCCACCTTCTTCGGTTTGTGGCCCGCCGGTTAG
ATTCCCGGGCCCGAGGACTACGAAGGCGCTGAAATCAGCGACGGCGCAGGAGAACTTGATTCT
TCAGTCCGCATAGCTGGTGGCCGATCATGGTCGCGTTGTCCGGCTCGGTGGCAGCGGTGGCA
15 TCGCGTTGTGGCTCCCGTGCTGATCGCCGCCGGTGTGGCATTATCCTCGCCTCGGCGGCCG
GATTGGTCTTCGAATATTACGTCGGTCCTGAGAAGCACTGA

>Rv2200c ctaC TB.seq 2463661:2464749 MW:40449

>emb|AL123456|MTBH37RV:c2464749-2463658, ctaC SEQ ID NO:90

20 GTGACACCTCGCGGGCCAGGTCGTTTGCAACGCTTGTGCGAGTGCAGGCCTCAGCGCGGCTCC
GGAGGGCCTGCCCCTGGTCTTCGACAGCTGGCGCTCGCAGCAATGCTGGGGCATTGGCCGT
CACCGTCAGTGGATGCAGCTGGTCGGAAGCCCTGGGCATCGGTTGGCCGGAGGGCATTACCC
CGGAGGCACACCTCAATCGAGAACTGTGGATCGGGGCGGTGATCGCCTCCCTGGCGGTTGGG
GTAATCGTGTGGGTCTCATCTTCTGGTCCGCGGTATTTACCGGAAGAAGAACCCGACACTG
25 AGTTGCCCCGCCAGTTCGGCTACAACATGCCGCTAGAGCTGGTTCTCACCGTCATACCGTTCT
CATCATCTCGGTGCTGTTTTATTTACCGTCGTGGTGCAGGAGAAGATGCTGCAGATAGCCAAG
GATCCCAGAGTCTGATTGATATCACGTCTTTCCAGTGAATTGGAAGTTTGGCTATCAAAGGGT
GAACTTCAAAGACGGCACACTGACCTATGATGGTGCCGATCCGGAGCGCAAGCGCGCCATGGT
TTCCAAGCCAGAGGGCAAGGACAAGTACGGCGAAGAGCTGGTCCGGCCGGTGCGCGGGCTCA
30 ACACCGAGGACCGGACCTACCTGAATTTGACAAGGTCGAGACGTTGGGCACCAGCACCGAAA
TTCCGGTGCTGGTGCTGCCGTCCGGCAAGCGTATCGAATTCCAAATGGCCTCAGCCGATGTGAT
ACACGCATTCTGGGTGCCGGAGTTCTTGTTCAAGCGTGACGTGATGCCTAACCCGGTGGCAAAC
AACTCGGTCAACGTCTTCAGATCGAAGAAATACCAAGACCGGAGCATTGCTGGGCCACTGCG
CCGAGATGTGTGGCACGTATCACTCGATGATGAACTTCGAGGTCCGCGTCGTGACCCCCAACG
35 ATTTCAAGGCCTACCTGCAGCAACGCATCGACGGGAAGACAAACGCCGAGGCCCTGCGGGCGA
TCAACCAGCCGCCCTTGCGGTGACCACCCACCGTTTGATACTCGCCGCGGTGAATTGGCCC
CGCAGCCCGTAGGTTAG

>Rv2427c proA g-glutamyl phosphate reductase TB.seq 2724231:2725475 MW:43746

>emb|AL123456|MTBH37RV:c2725475-2724228, proA SEQ ID NO:91

ATGACCGTGCCAGCACCGTCGCAGCTCGACTTGCGTCAAGAGGTGCACGACGCCGCACGCCG
5 CGCCCGGGTGGCCGCCCGCGGCTGGCATCGCTGCCGACGACTGTCAAAGACCGCGCGCTGC
ACGCGGCTGCCGACGAGCTACTGGCTCACC GCGACCAGATCCTGGCGGCCAACGCCGAAGAC
CTGAACGCGGCGCGCGAGGCGGACACCCCGCCGCCATGCTGGACCGGTTGTCTTGAACCC
GCAACGAGTCGACGGTATCGCCGCCGGGTTGCGGCAAGTCGCGGGACTGCGCGATCCGGTCG
GTGAAGTGCTGCGTGGCTATACCCTGCCAACGGGCTGCAGCTGCGCCAGCAGCGCGTCCCCC
10 TGGGCGTGGTCGGCATGATCTACGAGGGCCGCCCAATGTCACCGTGGATGCCTTCGGGCTGA
CACTCAAGTCGGTAACGCTGCATTGCTGCGCGGCAGCTCGTCGGCCGCAAAGTCCAACGAGG
CCCTGGTGGCGGTGTTACGCACCGCGCTGGTCGGCCTGGAGCTGCCGGCCGACGCGGTCCAG
CTGCTGTGGCTGCCGACCGCGCCACCGTCACTCACCTGATTAGGCCCGCGGCCTGGTCGAT
GTGGTGATTCCACGCGGGGGAGCGGGCCTGATCGAGGCGGTGCTACGCGATGCCAGGTGCC
15 CACCATCGAGACCGGCGTCGGGAACTGCCATGTCTACGTGCACCAAGCGGCCGACCTGGACGT
GGCCGAGCGTATCTTGCTGAACTCCAAGACGCGCGGCCAGCGTCTGCAACGCCGCCGAGA
CGCTGCTGGTCGACGCGAGCGATCGCCGAAACGGCGTTGCCTCGATTGCTGGCCGCCCTGCAGC
ACGCCGGTGTACCGTACATCTCGACCCGGACGAGGCCGACCTGCGCCGCGAATACCTGTCGC
TGGACATCGCGGTGGCGGTGGTCGACGGTGTGACGCTGCCATCGCCATATCAACGAATACG
20 GCACCGGGCACACAGAAGCGATTGTGACCACCAATCTTGATGCGGCCAACGCTTTACCGAACA
GATCGATGCGGCCGCGGTGATGGTGAACGCATCAACGGCGTTACCGACGGCGAGCAATTGCG
CTTCGGCGCCGAGATCGGCATCTCCACCCAGAACTGCATGCCCGCGGACCGATGGGACTACC
GGAATTGACGTCGACCAAGTGGATGCGATGGGGAGCCGGCCACACCCGTCCGGCCTGA

25 >Rv2438c - similar to YHN4_YEAST P38795 TB.seq 2734793:2737006 MW:80492

>emb|AL123456|MTBH37RV:c2737006-2734790, Rv2438c SEQ ID NO:92

ATGGGACTGCTCGGCGGCCAATCAGGGCCCAGGGTCGGCAGCGGCCAGTCGGTAGCATCCC
CACGCCGGTCAATGCCGCCATCTGCCAGCAGCGCGGGGGATTCCACGGTGTGAGCGTGGAT
ACTCGGCGGGTGATTGCGGCGTTCTGACGTCGCTGGGCGACAATGAAAGGACGATGAACTTTT
30 ACTCCGCCTACCAGCACGGGTTGCTGCGCGTTGCCGCTGCACTCACCACACCACCATCGGTG
ACCCGGCGGCCAACGCCGCGTCGGTATTGGACATGGCCCGTGCGTGCCACGACGATGGCGCA
GCGTTGGCGGTCTTTCCTGAGCTGACGCTGTGCGGGCTACTCCATCGAGGACGTACTACTGCAG
GACTCTCTGCTCGATGCCGTGAGGACGCGTGTGCTGACCTGGTGACCGAATCCGCCGACCTG
TTACCTGTACTGGTGGTCGGGGCTCCGCTGCGGCATCGACACCGCATCTACAACACCGCGGTC
35 GTCATTACCGCGGCGCCGTGCTCGGCGTGGTGCCCAAGTCGTATCTACCCACCTATCGCGAG
TTCTACGAGCGGCGCCAGATGGCGCCCGGAGACGGGGAGCGGGGCACGATCCGCATCGGTGG
CGCCGACGTGGCCTTCGGCACGGACCTGTTGTTGCGCGCTCAGATCTACCCGGCTTTGTGTT

GCATGTGGAGATCTGCGAGGACATGTTTGTGCCGATGCCGCCAGCGCCGAGGCGGCCCTGG
CGGGCGCGACGGTGCTGGCGAATCTGTCCGGCAGCCCGATCACCATCGGCCGTGCCGAGGAC
CGCCGGCTGCTTGCGCGCTCGGCGTCGGCGCGGTGTCTGGCTGCCTATGTCTATGCCGCCGC
GGGGGAGGGGAGTCAACGACGGACCTGGCCTGGGACGGTCAGACGATGATCTGGGAGAATG
5 GCGCACTGCTCGCGGAGTCCGAACGTTTCCCCAAAGGAGTGCGCCGCAGTGTGCCCGACGTTG
ACACCGAGTTGCTTCGGTCGGAGCGGCTGCGGATGGGCACGTTGACGACAACCGGCGTCAC
CACCGGGAGTTAACGGAATCGTTCCGGCGCATCGACTTCGCACTCGACCCACCGGCAGGCGAC
ATCGGACTGCTGCGCGAGGTGAGCGGTTCCCGTTTCGTTCCGGCCGATCCGCAACGATTGCAA
CAGGATTGCTACGAGGCCTACAACATCCAGGTGTCTGGACTCGAGCAACGGTTGCGGGCGCTG
10 GACTATCCGAAGGTGCTTATCGGTGTGTCCGGGGGATTGGAATCGACGCACGCGCTGATCGTC
GCGACCCATGCCATGGACCGCGAGGGCCGGCCGCGCAGCGACATTCTGGCGTTTGCGTTGCC
CGGATTGCCACCGGGGAGCACACTAAGAACAACCGGATCAAGCTGGCACGTGCGCTGGGGG
TTACCTTCTCCGAAATCGATATCGGCGACACCGCTCGGTTGATGCTGCACACAATCGGCCATCC
GTATTCGGTTGGCGAAAAAGTGACGACGTACCTTCGAGAACGTCCAGGCCGGGTGCGCAC
15 CGACTATCTTTCCGTATCGCCAACCAGCGCGGGGGAATCGTACTGGGCACCGGGGACCTGTC
GGAGCTGGCACTGGGTTGGTCGACATACGGTGTCCGGCACCAGATGTCCCACTACAACGTCAA
CGCCGGTGTGCCAAGACGCTGATCCAGCACCTGATCCGGTGGGTCAATTCGGCGGGTGAGTT
CGGTGAGAAGGTGGGTGAGGTATTGCAGTCGGTGCTCGACACCGAGATCACCCCCGAATCAT
TCCGACCGGCGAGGAGGAGCTGCAGAGCAGCGAGGCCAAGGTGCGACCTTTCGCCCTACAGG
20 ACTTTTCGCTTTTTCAGGTAATGCGCTACGGATTCGCCCGTCGAAGATTGCGTTTTTGGCCTGG
CATGCGTGGAACGATGCGGAGCGGGGCAACTGGCCGCCCGGCTTCCCAAAGAGCGAACGCCC
GTCCTATTCAATTGGCCGAAATCCGGCATTGGCTGCAGATTTTCGTCCAGCGGTTTATTGTTTA
GCCAGTTCAAGCGTTCCGCATTGCCCAACGGCCCCAAGGTGTCCACGGGGGCGCGTTGTCCG
CGCGTGGGGATTGGCGGGCCCCGTGCGATATGTCAGCGCGAATCTGGCTCGATCAGATCGACC
25 GTGAGGTGCCCAAGGGCTAG

>Rv2439c proB glutamate 5-kinase TB.seq 2737118:2738245 MW:38789

>emb|AL123456|MTBH37RV:c2738245-2737115, proB SEQ ID NO:93

ATGAGAAGTCCGCATCGGGACGCAATCCGGACCGCGCGCGCCCTTGTCTGAAGGTGCGGAC
30 CACGGCGCTTACCACACCGTCCGGGATGTTGATGCCGGCCGGCTGGCCGGAATGGCCGAGG
CGGTGAGCGGCGGATGAAGGCGGGTCCGACGTGTCATCGTGTCTTCGGGCGCCATCGCC
GCCGGCATCGAGCCGCTCGGGCTGTCCCGTCTCCCAAAGATCTGGCGACCAAGCAGGCGGC
GGCCAGCGTCGGGCAGGTGCGCTGGTGAATCGTGAGCGCGGCGTTGCCCGCTACGGCC
GCACGGTGGGCCAGGTGCTGCTGACCGCGCACGACATTTGATGCGGGTGCAGCACACCAAC
35 GCCCAACGCACGCTGGATCGGCTGCGCGCGTTGCACGCGGTGGCGATTGTCAACGAGAACGA
CACCGTGGCCACCAACGAGATCCGGTTCGGTGACAACGATCGGCTGTCTGCACTGGTGGCGCA
CCTGGTGGCGCCGACGCTTTGGTGCTGCTGTGCGACATCGACGGCCTCTACGACTGCGACCC

GCGCAAAACCGCGGACGCGACGTTTCATTCCGGAGGTGTCCGGGCCGGCGGATCTGGACGGTG
TGGTCGCCCGGCCGAGTAGCCACCTGGGTACTGGCGGCATGGCGTCCAAGGTGGCGGCCGGCG
CTGTTGGCCGCCGACGCCGGGTGCCGGTACTGCTGGCCCCGCGGCCGACGCCGCGACCG
CGCTCGCCGACGCGTGGTGGGCACGGTGTTTTCGGGCCCGGCCGCGGTCTGTGGCCCCGG
5 CGGTTCTGGGTGCGTTATGCCGCCGAAGCAACCGGCGCACTGACTCTCGACGCCGGTGCGGTG
CGCGCTGTGGTGCACAACGCCGGTCACTGCTGGCGGCCGGTATCACCGCGGTGTCCGGCCG
GTTTTGCGGCCGGCGATGTGGTCAACTGCGTGCACCCGACGCGGCCATGGTAGCCCCGCGGGG
TGGTTGCCACGACGCGTCCGAGCTGGCCACCATGGTGGGCCGGTCCACCTCTGAGCTACCCG
GCGAGCTGCGCCGCCCGGTGGTGCACGCCGACGATCTGGTGGCGGTGTGGCGAAGCAAGCT
10 AAGCAAGTTTAG

>Rv2440c obg Obg GTP-binding protein TB.seq 2738248:2739684 MW:50430

>emb|AL123456|MTBH37RV:c2739684-2738245, obg SEQ ID NO:94

GTGCCTCGGTTTGTGATCGGGTCGTATCCACACAGAGCGGGTTCGGGCGGTAACGGCTGC
15 GCTTCGGTCCATCGCGAGAAATCAAGCCGCTGGCGGCCCGGATGGCGGAAATGGCGGCCG
GGGCGGCAGCATCGTCTTCGTGTCGATCCGCAAGTGACACCCCTGCTCGACTTCCATTTCCG
CCGCATCTACCCGCGGCTTCGGGCAAGCACGGGATGGGCAATAACCGCGACGGGGCCGCCGG
CGCGGATTTGGAAGTGAAAGTTCCCGAAGGCACCGTGGTATTGGACGAGAACGGCCGGCTACT
GGCCGACCTGGTGGCGCGGGCACCCGCTTTGAAGCCGCCCGCGAGGCCGTGGCGGTTTGG
20 GCAACGCCGCGCTGGCTTCCCGCGTGCCTAAGGCCCGGTTTCGCACTCCTCGGCGAAAAGG
GACAGTCCCGAGACCTCACCTTGAACTCAAGACCGTCGCCGACGTCGGCCTGGTGGGTTTC
CGTCGGCCGAAAATCCTCGCTGGTGTGCGCGATTTGGCGGGCCAAGCCGAAGATCGCCGACT
ACCCGTTACCAACCTGGTGCCAACCTCGGTGTGGTCTCGGCTGGCGAGCACGCGTTCACCG
TCGCCGACGTGCCGGGTTGATCCCGGGCGCATCCCGGGGCCGTGGTCTGGGGCTGGACTTT
25 CTGCGGCACATCGAGCGCTGCGCTGTAAGTGGTGCATGTGGTGGATTGCGCTACCGCCGAGCCG
GGCCGCGACCCCATCTCGGACATCGACGCGCTGGAAACGGAACCTCGCGTGCTACACGCCAC
GCTGCAAGGGGACGCGGCTCTGGGCGATCTCGCCGACGGCCGCGTGGGTGGTCTCAACA
AAATCGATGTGCCGAGGCCCGCGAGCTCGCGGAGTTCGTCCGTGACGACATCGCCAGCGC
GGCTGGCCGGTGTCTGCGTGTGACCGCAACCCGGGAAAACCTGCAGCCGTTGATCTTTGGG
30 CTGTGCGCAGATGATCTCGGACTACAACGCTGCGCGGCCGGTGGCGGTGCCACGGCGGCCGGT
GATTCGTCCGATTCGGTGGACGACAGCGTTTTACCGTCGAACCCGACGGGCATGGTGGCTT
TGTCGTCAGCGGTGCCCGGCCCGAGCGTTGGATTGACCAGACCAACTTCGACAACGACGAGGC
CGTCGGCTATCTCGCCGACCGGCTGGCGCGCTGGGTGTGAGGAGGAATTGCTGAGGCTGG
GTGCGCGGTGAGGATGCGCGGTGACCATCGGCGAGATGACGTTGATTGGGAGCCGCAAACG
35 CCTGCGGGTGAGCCGGTGCAGATGTCCGGCCGGGGCACCGATCCGCGGCTGGACAGCAACAA
GCGGGTGGGCGCGGCCGAGCGAAAGGCCGCTCGGAGTCGGCGTCCGAACACGGGGATGGC
TGA

>Rv2441c rpmA 50S ribosomal protein L27 TB.seq 2739773:2740030 MW:8969

>emb|AL123456|MTBH37RV:c2740030-2739770, rpmA SEQ ID NO:95

ATGGCACACAAGAAGGGGGCTTCCAGCTCGCGCAACGGTCGCGATTCCGCCGCCAGCGGCT
5 GGGGGTTAAGCGGTACGGCGGCCAGGTCGTCAAGGCCGGCGAGATCCTGGTCCGCCAGCGCG
GTACCAAATTCATCCCGGCGTCAACGTCGGGCGTGGCGGCGATGACACCTTGTTCCGCAAGA
CGGCCGGGGCGGTCTGAGTTCGGCATCAAACGCGGACGTAAGACGGTGAGCATCGTCGGTTCC
ACCACTGCCTGA

10 >Rv2442c rplU 50S ribosomal protein L21 TB.seq 2740048:2740359 MW:11152

>emb|AL123456|MTBH37RV:c2740359-2740045, rplU SEQ ID NO:96

ATGATGGCGACCTACGCAATCGTCAAGACCGGCGCAAGCAGTACAAAGTCGCTGTCGGAGAT
GTGGTCAAGGTGCAAAAGCTGGAATCCGAGCAGGGGAGAAGGTGTCCCTGCCGGTGGCTCT
GGTTGTGCACGGCGCCACCGTCACCACCGATGCGAAGGCACTGGCCAAGGTGCGGGTGACCG
15 GTGAGGTGCTCGGGCACACCAAGGGCCCCAAGATCCGTATCCACAAGTTCAAGAACAAAGACTG
GCTACCACAAACGGCAGGGACACCGTCAGCAGCTGACGGTCCTGAAGGTCACCGGCATCGCAT
AA

>Rv2448c valS valyl-tRNA synthase TB.seq 2747596:2750223 MW:97822

20 >emb|AL123456|MTBH37RV:c2750223-2747593, valS SEQ ID NO:97

ATGCTGCCCAAGTCGTGGGATCCGGCCGCGATGGAGAGCGCCATCTATCAGAAAGTGCTGGAC
GCTGGCTACTTCACCGCGGACCCGACCAGCACCAAGCCGGCCTATTCGATCGTGCTGCCGCCG
CCGAACGTGACCGGCAGCCTGCACATGGGCCACGCGCTGGAACACACCATGATGGACGCCTTG
ACGCGGCGCAAGCGGATGCAGGGCTATGAGGTGCTCTGGCAGCCGGGCACCGACCATGCCGG
25 GATCGCCACCCAGAGCGTGGTCGAGCAGCAGCTGGCGGTGACGGCAAGACTAAAGAAGACCT
CGGCCGCGAGCTGTTCTGTGACAAGGTGTGGGATTGGAAGCGAGAGTCTGGCGGTGCCATCG
GCGGCCAGATGCGCCGACTCGGTGACGGGGTGGACTGGAGCCGCGACCGGTTACCATGGAC
GAAGGTCTGTCGCGGGCGGTGCGCACGATCTTCAAGCGGCTTTATGACGCCGGGCTGATCTAT
CGGGCCGAGCGGCTGGTCAACTGGTCGCCGGTGCTGCAGACCGCGATCTCCGACCTCGAGGT
30 CAACTACCGCGACGTGCAAGGCGAGCTGGTGTGCTTTAGGTACGGCTCGCTTGACGACTCGCA
ACCCACATCGTGGTCGCCACCACCCGGTTCGAGACGATGCTGGGCGATACCGCGATCGCCGT
CCATCCCGATGACGAGCGCTACCGTCACCTGGTCGGCACCGCCTGGCGCACCCATTCTGTCGA
CCGGGAGCTGGCCATTGTCGCCGACGAGCACGTGGACCCTGAATTCGGCACCGGCGCGGTCA
AAGTCACACCCGCCCACGACCCCAACGACTTCGAAATCGGGGTGCGCCACCAGCTGCCGATGC
35 CCTCGATCCTGGACACCAAGGGCCGGATCGTCGACACCGGAACGCGATTGACGGCATGGACC
GCTTCGAGGCACGGGTGCGGTGCGCCAAGCGCTCGCGGCCAGGGCCGCGTGGTTCGAAGAA
AAGCGACCCTACCTGCACAGCGTCGGACACTCCGAACGCAGCGGCGAGCCGATCGAGCCGCG

GCTATCCCTGCAGTGGTGGGTCCGGGTGGAATCGCTGGCCAAAGCGGCCGGGGATGCGGTGC
GCAACGGGGACACCGTGATTACCCCGCCAGCATGGAACCCCGCTGGTTCTCCTGGGTGCACG
ACATGCACGACTGGTGCATCTCGCGACAGCTCTGGTGGGGGCATCGGATCCCGATCTGGTACG
GACCCGACGGCGAACAGGTGTGCGTCGGCCCGGACGAAACACCCCGCAGGGCTGGGAACAG
5 GATCCTGACGTGCTGGATACCTGGTTTTCGTCGGCGCTGTGGCCGTTTTCCACGCTGGGTGGC
CGGACAAGACGGCGGAGCTGGAAAAGTTCTATCCGACAAGCGTTCTGGTTACCGGCTATGACAT
CTTGTTCTTTGGGTGGCCAGAATGATGATGTTCCGGCACCTTCGTGGCGACGACGCCGCCATC
ACCCTCGACGGCCGCCGGGGCCCGCAGGTGCCGTTACCGACGTGTTTCTGCATGGGCTGATC
CGCGACGAGTCTGGCCGCAAGATGAGCAAGTCCAAGGGCAACGTCATCGACCCGCTGGATTGG
10 GTGGAAATGTTCCGGGGCCGATGCGCTGCGGTTACGCTGGCCCGCGGGGCCAGTCCCGGTGG
TGACTTGCGGTGAGCGAGGATGCCGTGCGGGCGTCGCGCAATTTGGGACCAAGCTGTTCAA
CGCCACTCGGTACGCACTGCTCAATGGCGCCGCGCCAGCACCCCTGCCATCGCCGAACGAGCT
GACCGACGCCGACCGCTGGATTCTCGGAAGGTTGGAAGAGGTTCCGGGCCGAAGTTGATTGGC
CTTCGACGGATACGAGTTCAGCCGCGCTTGTGAGTCCCTGTATCACTTCGCTGGGACGAATTC
15 TCGGACTGGTACCTCGAACTGGCCAAAACGCAGCTTGCCAGGGACTCACACACACCACCGCC
GTGCTGGCCGCCGGGCTGGACACGCTGCTGCGCCTGCTGCACCCGGTGATTCCCTTCCTCACC
GAGGCGCTATGGCTGGCGCTGACCGGCAGGGAATCGCTGGTCAGCGCCGACTGGCCGGAGCC
TTCCGGGATTAGCGTGGACCTTGTTGCCGCGCAACGGATTAACGATATGCAGAAGTTGGTGACC
GAAGTGCGGCGGTTCCGCAGCGATCAAGGTCTGGCCGACCGGCAGAAGGTTCCGGCCCCGAAT
20 GCACGCTGTGCGGGACTCGGATCTGAGCAACCAGGTGGCCGCCGTGACCTCGCTGGCGTGGC
TCACCGAGCCGGGCCCGGATTTGAGCCGTGGTCTCGTTGGAGGTTCCGGCTCGGCCCCGAGA
TGAACCGCACCGTCGTGCTGAGCTCGACACCTCGGGCACCATCGACGTGGCCGCCGAGCGT
CGCCGCTGGAAGGAGTTGGCCGGCGCCAAAAGGAGCTGGCGTCGACCGCCGCCAAGTT
GGCCAACGCGGACTTTCTGGCCAAAGCGCCGACGCCGTATTGCCAAGATCCGGGACCGCCA
25 GCGCGTGGCGCAGCAGGAAACCGAGCGCATCACCACCCGTTGGCTGCGCTGCAATGA

>Rv2482c plsB2 TB.seq 2786915:2789281 MW:88284 >emb|AL123456|MTBH37RV:c2789281-
2786912, plsB2 SEQ ID NO:98

GTGACCAAACCGGCGGCCGATGCCAGCGCGGTGCTTACTGCCGAGGACACACTGGTGTGGC
30 TTCCACGGCGACGCCGCTGAGATGGAGCTGATCATGGGCTGGCTGGGCCAGCAGCGTGAC
GCCATCCGGACTCGAAGTTCGACATATTGAAGCTGCCACCGCGCAACGCTCCGCCGGCGGCGC
TGACGGCACTGGTCGAGCAGCTCGAGCCCGGCTTCGCATCCAGCCCGCAATCTGGCGAGGAC
CGTTCTATCGTGCCGTTTCGGGTGATCTGGCTGCCTCCCGCCGATCGCAGCCGGGCGGGCAAG
GTGGCCGCACTGCTCCCGGGTCGGGATCCCTACCATCCCAGCCAGCGTCAGCAGCGTCGCATC
35 CTGCGTACCGATCCCAGGCGCGCGGGTGGTGGCCGGCGAGTCGGCCAAGGTGTCCGAAC
GCGCCAGCAGTGGCGCGATACACGGTGGCAGAGCACAAGCGCGATTTCGCCAGTTCTGTCAG
CCGCCGAGCGCTGTTGGCGCTGGCGCGCGCGAATATCGGATCCTTGACCGCAATACAAATC

TCCCCGGCTGGTGAAGCCGGAGATGTTGGCGTCCGCACGATTCGTGCCGGCCTGGACCGGAT
TCCGGGCGCCACGGTCAAGATGCCGGGAAGATGCTCGACGAACTCTCCACCGGATGGAGCC
AGGTGTCGGTAGACCTGGTTTCCGTCTCTCGGCAGGCTGGCTAGCCGCGGCTTCGATCCGGAAT
TCGACTACGACGAGTATCAGGTGCGCGCGATGCGCGCCGCACTGGAGGCTCATCCGGCGGTC
5 CTGCTGTTCTCGCACCGGTCTACATCGACGGCGTGGTGGTACCGGTGGCCATGCAGGACAAC
CGGTTACCGCCGGTGACATGTTCCGGCGGCATCAACCTGTCGTTCCGTCTCATGGGACCCCTC
ATGCGGCGCTCGGGGATGATCTTCATCCGGCGCAATATCGGCAACGACCCACTGTATAAGTACG
TGCTCAAGGAGTACGTGGGCTACGTGGTTCGAGAAGCGGTTCAACCTGAGCTGGTCCATCGAAG
GCACCCGGTTCGCGCACCGGAAAGATGTTGCCGCCAAGCTCGGTTTGATGAGCTACGTGGCCG
10 ATGCTTACCTGGACGGCCGCACTGACGACATCCTGCTGCAGGGGGTTTCGATTTGCTTCGATCA
GCTGCACGAGATCACCGAATACGCCGCTACGCGCGTGGCGCGGAGAAGACGCCCGAAGGTT
TGCGCTGGCTCTACAACTTCATCAAGGCGCAGGGGGAACGCAACTTCGGCAAGATCTACGTTCCG
CTTCCCCGAAGCGGTCTCGATGCGCCAGTACCTCGGCGCACCGCGAGCTGACCCAGG
ATCCGGCCGCGAAACGGCTTGCGTTGCAGAAGATGTCGTTGAGGTGGCCTGGAGGATTTTGC
15 AGGCGACGCCGGTGACCGCGACGGGTTTGGTGTCCGCACTGCTGCTCACCACCCGCGGCACC
GCGTTGACGCTCGACCAGCTGCACCACAGTTGCAGGACTCACTGGACTATCTGGAACGCAAA
CAATCGCCGGTTTCGACAAGCGCATTGCGACTGCGCTCGCGCAAGGCGTCCGTGCGGCGGC
GGACGCGTTGTCCAACGGCCACCCGGTCACTCGGGTGCACAGTGGCCGGGAGCCGGTATGGT
ACATAGCGCCTGACGACGAGCACGCCGCGGCGTTCTACCGGAACTCGGTGATCCATGCGTTTTT
20 GGAGACCTCGATCGTTCGAGCTCGCGCTGGCCCATGCCAAGCACGCCGAAGGTGACCGCGTCG
CCGCGTTCTGGGCCAGGCGATGCGGTTGCGGGATCTGCTGAAGTTGCACTTCTATTTGCGCG
ATTCCACGGCGTTTCGGGCCAACATCGCCCAAGAGATGGCCTGGCACCAAGACTGGGAGGATC
ATCTTGGCGTCGGGGGCAATGAGATCGACGCGATGCTGTATGCCAAACGGCCGCTGATGTCGG
ACGCGATGTTGCGGGTCTTCTTGAAGCCTATGAGATCGTTGCCGACGTGTTGCGCGATGCTCC
25 GCCTGACATCGGTCTTGAGGAGTTGACGGAGCTGGCGCTCGGCCTCGGCCGTCAGTTTGTGGC
ACAGGGCCGGGTCCGCAGCAGCGAACCAGGTATCGACGCTGCTGTTCCGCACTGCACGCCAGG
TCGCCGTCGATCAGGAGCTGATAGCGCCGGCGGCCGACCTCGCCGAACGTAGGGTGCCTTC
CGGCGGGAGTTACGAAACATTCTGCGGGATTTGCACTATGTGAGCAGATCGCGCGCAACCAG
TTCGTGCGCTGCGAGTTCAAAGCGCGTCAAGGACGCGACCGAATCTAA

30

>Rv2509 - putative oxidoreductase TB.seq 2824676:2825479 MW:28014

>emb|AL123456|MTBH37RV:2824676-2825482, Rv2509 SEQ ID NO:99

ATGCCGATACCCGCGCCAGCCCCGACGCACGTGCCGTTGTCACCGGGGCTTCGCAGAACATC
GGCGCGGCGCTGGCCACCGAACTGGCCGCACGCGGGCACCACTGATCGTCACCGCACGACG
35 CGAGGACGTGTTGACCGAGTTGGCTGCCCGGCTGGCCGACAAGTACCGCGTCACGGTCGACG
TGCGACCGGCCGATCTGGCCGATCCGCAAGAACGATCGAACTGGCCGACGAGCTGGCTGCC
CGGCCCATCTCGATCCTGTGCGCCAACGCGGGTACCGCGACATTGGGCCGATCGCATCGCTC

GATCTTGCCGGCGAAAAGACGCAGGTGCAGTTGAATGCCGTGGCGGTGCACGACCTTACGTTG
GCGGTGTTGCCGGGCATGATCGAGCGCAAGGCCGGCGGCATCTTGATTTCTGGTTCGGCGGCC
GGCAATTCACCGATTCCCTACAACGCCACCTATGCCGCGACCAAGGCCTTCGTGAACACCTTCA
GCGAATCTCTGCGCGGTGAGCTACGCGGCTCCGGCGTGCACGTACGGTGCTGGCCCCGGGC
5 CCGGTTTCGACCGAGCTACCGGATGCCTCCGAAGCGTCACTGGTCGAGAAGCTGGTGCCGGAC
TTCTGTGGATCTCGACGGAGCACACCGCCCGGGTATCGCTGAATGCCTTGGAGCGCAACAAG
ATGCGCGTCTTCCGGGTCTGACGTCAAAGCGATGTCGGTGGCCAGCCAATACGCTCCGCGC
GCCATCGTGGCGCCAATCGTGGGTGCCTTTACAAGAGGCTTGGGGGCAGCTAG

10 >Rv2524c fas fatty acid synthase TB.seq 2840124:2849330 MW:326226

>emb|AL123456|MTBH37RV:c2849330-2840121, fas SEQ ID NO:100

GTGACGATCCACGAGCACGACCGGGTGTCCGCTGATCGCGCGGGGACAGCCCGCATACCAC
CCACGCTCTGGTGCATCGCCTCATGGCTGGTGAGCCCTACGCTGTCGCATTCGGTGGCCAGGG
CAGCGCCTGGCTGGAACCCTCGAAGAGCTGGTGTCGGCCACCGGGATAGAAACCGAGTTGGC
15 GACGTTGGTCGGTGAGGCAGAGCTGTTGCTCGATCCGGTCACCGACGAGCTGATTGTGGTGCG
CCCGATCGGTTTCGAGCCGCTGCAATGGGTACGCGCACTGGCGGCCGAGGACCCGGTTCGGT
CCGACAAGCACCTGACGTGCGCCGCGGTGTCGGTGGCCGGCGTGTGCTTACCCAGATCGCGG
CGACCCGGGCGCTGGCCCGTCAAGGCATGGACCTCGTGGCCACCCGCGCGGTGCGCATGGCG
GGGCATTGCAAGGTGTGCTGGCGGTGGAAGCCCTCAAGGCTGGTGGGGCACGCGACGTGCA
20 GCTGTTTGCTTGGCCCCAGTTGATCGGTGCCGCCGGAACGCTGGTGGCCCGCCGGCGCGGAA
TTTCCGTCTGGGCGATCGCCCGCCGATGGTATCGGTCACCAACGCCGACCCCGAGCGCATCG
GCCGGTTGCTCGACGAGTTCGCCCAGGACGTGCGCACGGTGCTGCCACCGGTGTTGTCCATCC
GCAACGGCCGCGCTGCCGTGCTCATCACCGGCACCCCGAGCAGCTGTCGCGTTTCGAGCTTT
ATTGCCGCCAGATCTCCGAGAAGGAAGAAGCCGACCGCAAGAACAAGGTCCGCGGCGGCGAC
25 GTCTTCTCGCCGGTCTTCGAGCCGGTGCAGGTGGAGGTGGGCTTTCACACCCCGCGGCTATCC
GACGGGATCGACATCGTCGCGGGCTGGGCCGAGAAGGCGGGCCTCGATGTGCGCTTGGCTCG
GGAGCTGGCCGATGCCATCTTGATCAGAAAGGTGCACTGGGTGACGAGATCACCCGTGTCCA
CGCGGCCGGCGCCCGCTGGATCCTCGACCTGGGGCCGGGCGACATCCTGACCCGACTGACCG
CACCGGTGATCCGCGGCCTGGGCATCGGCATCGTGCCGGCGGCTACCCGCGGTGGCCAGCGC
30 AACCTGTTACCGTGGCGCCACCCCGAGGTTGCCCGGGCCTGGTCGAGCTACGCACCGACC
GTGGTTGCGCTCCCCGACGGCAGGGTCAAGCTCTCGACGAAGTTCACCCGGCTGACCGGCCGC
TCGCCGATCCTGCTCGCGGGCATGACCCCGACACCGTGGACGCCAAGATCGTCGCCGCGGC
GGCCAACGCCGGGCACTGGGCCGAGCTGGCCGGCGGCGGGCAGGTCACCGAAGAGATCTTC
GGTAACCGCATCGAACAATGGCCGGCCTGCTCGAGCCGGGCGGCACCTATCAGTTCAACGCG
35 CTGTTCTCGATCCCTACCTGTGGAAGCTTCAGGTGGGCGGCAAGCGGTTGGTGCAGAAGGCC
CGCCAGTCCGGCGCCGCGATCGACGGCGTGGTGATCAGCGCCGGCATCCAGACCTCGACGA
GGCCGTGAGCTGATCGACGAAGTGGGCGACATCGGCATCAGCCACGTCGTGTTCAAACCCGG

GACCATCGAGCAGATCCGCTCGGTGATTGCGATCGCCACCGAGGTGCCCACCAAGCCGGTGAT
CATGCACGTCGAGGGCGGGCGCGCCGGCGGGCACCATTCCCTGGGAGGATCTCGACGACCTGC
TGCTGGCTACCTACTCGGAGTTGCGCTCACGCGCCAACATCACGGTGTGCGTCGGCGGGCGGCA
TTGGCACCCCGAGAAGGGGCTGCGGAATATTTGTCCGGGCGCTGGGCGCAGGCCTACGGCTTCC
5 CATTGATGCCGATCGACGGCATCCTGGTCGGCACC GCGGCGATGGCCACCAAGGAATCCACCA
CGTCGCCATCGGTCAAGCGGATGCTCGTCGACACTCAGGGCACCGACCAATGGATCAGCGCCG
GAAAAGCGCAGGGCGGCATGGCCTCCAGCCGCGAGTCAGCTCGGTGCCGATATCCACGAGATC
GACAACAGCGCATCCCGGTGCGGGCGGCTGCTCGACGAGGTGGCCGGTGACGCGGAGGCGG
TCGCGGAGCGTCGCGACGAGATCATCGCGGCGATGGCCAAGACCGCCAAGCCCTACTTCGGC
10 GACGTCGCCGACATGACCTACCTGCAGTGGCTGCGGCGCTACGTCGAACTGGCCATCGGGGAA
GGCAACTCGACCGCCGACACCGCCTCGGTGGGCAGCCCGTGGCTGGCCGACACCTGGCGGGA
CCGCTTCGAGCAGATGCTGCAGCGTGCCGAAGCCCGGTTGCACCCACAGGATTCGCCCCGAT
CCAGACGCTATTACCGATGCTGGCCTGCTGGACAATCCGCGAGCAGGCGATCGCCGCCCTGCT
GGCGCGCTACCCCGACGCCGAGACCGTGCGAGTTGCATCCCGCGGATGTGCCCTTTTTCGTGAC
15 GTTGTGCAAGACGCTGGGCAAGCCGGTCAACTTCGTGCCGGTGATCGACCAGGACGTGCGGC
GCTGGTGGCGCAGCGACTCGCTGTGGCAGGCCACGACGCCCGCTACGACGCCGATGCGGTG
TGCATCATTCCGGGCACCGCGTCGGTAGCCGGCATCACCCGGATGGATGAACCCGTGCGGTGAG
TTGCTGGACCGTTTCGAGCAAGCCGCAATCGATGAAGTGCTCGGCGCCGGTGTCGAGCCGAAG
GATGTGCGCTCGCGCCGGCTGGGCCGCGCCGACGTGGCCGGACCGTTGGCTGTGCTCCTCGA
20 CGCACCCGATGTGCGCTGGGCCGGTCGCACCGTGACCAACCCGGTGATCGGATCGCCGACC
CGGCCGAATGGCAGGTGCACGATGGACCCGAAAACCCGCGCGCCACACACTCATCCACCGGC
GCCCGGCTGCAGACGCACGGCGACGACGTGCGCTTGAGCGTGCCCGTCTCGGGCACCTGGGT
CGACATCCGATTACGTTGCCGGCCAACACCGTCGATGGCGGCACCCCGGTGATCGCCACCGA
GGACGCCACCAGCGCCATGCGCACGGTGCTGGCGATCGCCGCCGGTGTCGACAGCCCGGAGT
25 TCTTGCTGCGGTGGCCAACGGGACGGCCACTTTGACGGTGGACTGGCACCCCGAGCGTGTG
CCGACCACACCGGCGTCAACGCCACGTTCCGGTGAGCCGCTGGCACCCAGCCTCACCAACGTG
CCCGACGCGCTCGTCGGCCCTTGTGGCCAGCGGTTTTTCGCGGCCATCGGATCGGCGGTCAAC
GACACCGGTGAGCCGGTGGTGGAAGGCCTGCTGAGCCTGGTGATCTGGACCACGCCGCCCG
CGTGGTCGGTCAGCTGCCCACGGTCCCGGCCCAATTGACCGTCACCGCAACGGGTGCCAACGC
30 AACCGATACGGACATGGGCCGCGTCGTGCCGGTCTCGGTGCTCGTTACCGGCGCCGATGGCG
CCGTGATCGCCACTCTCGAGGAGCGATTGCGGATCCTGGGTGCGACCGGTTCCGCCGAGCTCG
CCGACCCGGCGCGAGCCGGTGGCGCGGTGTCGGCGAACGCCACCGACACCCCGCGCCGTGCG
CCGCCGCGACGTACGATCACCGCGCCGGTCGACATGCGCCCGTTGCGGGTGGTGTCGGGCG
ACCACAACCCCATTCACACCGACCGGGCCGCGCGCTGCTTGCCGGCCTGGAGTCGCCGATC
35 GTGCACGGCATGTGGCTGTGCGCCGCGGCGCAACACGCGGTGACCGCCACCGACGGGCGAGG
CCCGGCCACCGGCCCGGCTGGTGGCTGGACCGCGCGGTTTTTGGGCATGGTGCGCCCCGGC
GACGAGGTGGACTTCGCGCTCGAGCGCGTCGGAATCGACCAGGGCGCAGAGATTGTGGACGT

GGCCGCGCGCGTCGGGTCGGATCTAGTGATGTCGGCCTCCGCGCGACTGGCCGCACCCAAGA
CGGTCTACGCATTCCCCGGCCAGGGCATCCAACACAAGGGCATGGGCATGGAGGTGCGCGCC
CGCTCCAAGGCGGCCCCGCAAGGTGTGGGACACCGCGGACAAGTTCACCCGCGACACCCTGGG
CTTCTCGGTACTGCACGTGGTCCGCGACAACCCGACCAGCATCATCGCCAGCGGTGTGCACTA
5 CCACCACCCCGACGGGGTGCTCTACCTGACGCAGTTCACCCAGGTGCGCATGGCGACGGTGG
CGGCCGCGCAGGTGCGCCGAGATGCGTGAACAGGGAGCCTTCGTGGAAGGCGCCATCGCGTG
GGCCACTCGGTGCGCGAGTACACCGCGCTGGCCTGCGTGACCGGCATCTACCAACTGGAAGC
CTTGCTGGAGATGGTGTTTCACCGCGGGTGAAGATGCACGACATCGTTCCGCGCGACGAGCT
CGGCCGCTCCAACATATCGGCTGGCGGCCATCCGGCCGTCCCAGATCGACCTCGACGACGCCG
10 ACGTGCCCGCGTTCGTGCGCGGGATCGCGGAGAGCACCGGTGAATTCCTGGAGATCGTGAATT
TCAACCTGCGTGGCTCGCAATACCGCATCGCGGGCACGGTACGCGGCCTCGAGGCGCTCGAG
GCCGAGGTGGAGCGGCGCCGCGAGCTCACCGCGGCGCGACGGTCGTTCAATTTGGTGCCCGG
CATCGATGTTCCGTTCCACTCGCGAGTGCTGCGGGTGGGGTGGCCGAATTCGGGCGCTCGCT
GGACCGGGTCATGCCGCGCGACGCGGACCCCGACCTGATCATCGGGCGCTACATTCCCAACCT
15 GGTGCCGCGGTTGTTACCCCTGGACCGCGACTTCATCCAGGAAATCCGGGATTTGGTGCCCGC
CGAGCCGCTCGACGAGATCCTCGCCGACTACGACACCTGGCTTCGCGAGCGTCCGCGCGAGAT
GGCGCGCACGGTGTTTCATCGAGTGCTGGCATGGCAATTCGCCAGCCCGGTGCGCTGGATCGA
GACGCAGGATCTGCTGTTTCATCGAGGAGGCCGCCGGCGGGTGGGTGTGGAGCGATTGTCG
AGATCGGTGTGAAGAGCTACCGACGGTGGCGGGTCTTGCCACCAACACCCTCAAACCTGCCCG
20 AATACGCCCACAGCACAGTGAAGTGCTCAACGCCGAGCGTGATGCCGCGGTGCTGTTGCCCA
CCGACACCGACCCGGAGCCGGAGCCGGAGGAAGACGAGCCGGTCGCGGAATCGCCCGCGCC
GGACGTCGTCTCGGAAGCCGCCCCGTCGCGCCGGCCGCTTCGTGCGCGGGCCCCGCTCCCG
ACGATCTGGTTTTGACGCGCCCGCATGCCACGCTGGCGCTGATCGCGCTCTCGGCCAAGATGC
GCATCGACCAGATCGAAGAACTCGACTCCATCGAGTCCATCACCGACGGTGCGTCGTGCGGGC
25 GCAACCAGCTGCTGGTGGACCTGGGCTCCGAGCTGAACCTCGGTGCCATTGACGGCGCCGCC
GAATCGGACCTGGCCGGTCTGCGCTCACAGGTGACCAAACCTGGCGCGCACCTACAAGCCTTAC
GGCCCAGTGCTTTCCGACGCCATCAACGACCAGCTTCGCACCGTCTCGGACCGTCGGGCAAG
CGGCCCGGCGCCATCGCCGAGCGGGTGAAGAAGACCTGGGAGCTCGGTGAGGGCTGGGCCA
AGCATGTCACCGTCGAGGTGCGCTGGGCACCCGCGAGGGCAGCAGCGTTCGCGGGCGCGCC
30 ATGGGCCACCTGCACGAGGGCGCGCTGGCCGATGCCGCCTCCGTGACAAGGTCATCGACGC
GGCGGTGCGATCGGTGCGCGCGCGCCAGGGCGTTTCGGTAGCGCTGCCGTGCGCCGGTAGTG
GTGGCGGGCGCCACCATCGACGCGGGCCGCGCTCAGCGAGTTCACCGACCAATCACCGGCCGT
GAGGGCGTGCTGGCCTCCGCGGGCCCGCCTGGTGCTGGGGCAGCTGGGACTGGACGACCCCGT
CAACGCCTTGCCGGCCGCCCCGATTCCGAGCTGATCGACTTGCTCACCGCCGAACCTGGGAGC
35 GGA CTGGCCGCGGTTGGTGGCACCGGTGTTGACCCCAAGAAGGCCGTGCTATTGACGACC
GCTGGGCCAGCGCCCGCGAGGACCTGGTGAAGCTGTGGCTGACCGACGAGGGCGACATCGAC
GCCGACTGGCCGCGCCTGGCGGAGCGCTTCGAGGGTGCCGGCCACGTGCTGGCGACCCAGG

CTACCTGGTGGCAAGGTAAGTCGCTGGCCGCGGGCCGGCAGATCCATGCATCGCTGTACGGCC
GCATCGCCGCGCGCGCCGAGAACCCCGAACCCGGCCGCTACGGCGGCGAAGTTGCCGTGGTG
ACCGGCGCTTCGAAGGGTTCGATCGCCGCGTCGGTGGTGGCTCGGCTGCTCGACGGCGGAGC
CACCCTCATCGCGACCACCTCCAAGCTCGACGAGGAGCGGCTGGCGTTCTACCGCACGCTGTA
5 TCGCGACCACGCCCGTTACGGCGCGGCGCTGTGGCTGGTTCGCGGCGAACATGGCGTCCTACT
CCGACGTCGACGCCCTGGTCAATGGATCGGCACCGAACAGACCGAAAGCCTTGCGCGCGCAGT
CGATTACATCAAAGACGCGCAGACCCCGACGCTGCTGTTCCCGTTTCGCGGCGCCACGCGTGG
TCGGGGACCTGTTCGAGGGCCGGTTCGCGCGCCGAGATGGAGATGAAAGTGCTGCTGTGGGCC
GTGCAACGGCTGATCGGCGGCCTGTGACGATCGGCGCCGAACGCGACATCGCGTCGCGGCT
10 GCACGTGGTGTGCCCCGGCTCGCCCAACCGTGGCATGTTTCGCGGCGGACGGCGCCTACGGCG
AAGCCAAGTCCGCGCTGGATGCCGTGGTGAGCCGCTGGCAGCCGAGTCGTCCTGGGCGGCA
CGGGTCAGCCTGGCGCACGCGCTCATCGGCTGGACCCGCGGCACCGGGCTGATGGGCCACAA
CGATGCCATCGTGGCCGCCGTTCGAAGAGGCCGGGGTCAACACCTACTCGACCGACGAGATGG
CGGCGCTGCTGCTCGACCTGTGTGATGCGGAATCCAAGGTGGCTGCGGCGCGTTCGCCGATCA
15 AGGCCGACCTGACCGGGGGCCTGGCCGAGGCCAACCTCGACATGGCCGAGCTGGCGGCCAAG
GCGCGCGAGCAGATGTTCGGCAGCGGCGGCCGTTCGACGAGGACCGCGAGGCCCTGGCGCCA
TCGCCGCGCTGCCGTTCGCCGCCCGGGGTTTACCCCCGACCGCCGCGCAATGGGACGAC
CTCGATGTGACCCGCGCCGACCTGGTGGTGATCGTCGGCGGCGCCGAAATCGGCCCGTACGG
CTCGTCACGCACCCGGTTCGAGATGGAGGTGAAAAACGAGCTGTTCGGCGGCCGGCGTGCTGG
20 AGCTGGCCTGGACCACTGGGTTGATCCGCTGGGAGGACGACCCGCAACCCGGTTGGTACGACA
CCGAATCCGGCGAAATGGTTCGACGAATCCGAGTTGGTGCAGCGCTACCACGACGCCGTGGTGC
AGCGCGTCGGCATTTCGCGAATTCGTTGATGACGGCGCGATCGACCCCGACACGCCTCGCCGC
TGCTGGTGTTCGGTGTTCCTGGAGAAGGACTTCGCGTTTCGTGGTGTCTTCGGAGGCCGATGCGC
GCGCCTTCGTCGAGTTCGATCCCGAGCACACGGTCATCCGGCCGGTGCCCGACTCCACCGACT
25 GGCAGGTCATCCGCAAGGCCGGCACCGAGATCCGGGTGCCGCGAAAGACCAAGCTGTCCCGC
GTGCTCGGCGGCCAGATCCCGACCGGGTTCGACCCGACGGTGTGGGGCATCAGCGCAGACAT
GGCCGGTTCATCGACCGGTTGGCGGTATGGAACATGGTGGCGACCGTCGACGCGTTCTGTGTC
GTCCGGTTTCAGCCCGGCCGAGGTGATGCGTTACGTGCACCCGAGTTTGGTGGCCAACACCCA
GGGCACCGGCATGGGCGGCGGGCACGTGATGCAGACGATGTACCACGGCAATCTGTTGGGCC
30 GCAACAAGCCGAACGACATCTCCAGGAAGTCTTGCCGAATATCATTGCCGCGCACGTGGTTCA
GTCCTACGTCCGTAGCTACGGTGCGATGATCCACCCGGTAGCCGCGTGCGCCACCGCCGCGGT
GTCCGGTCGAGGAAGGTGTGACAAGATCCGGTTGGGCAAGGCTCAACTGGTGGTGGCCGGCG
GCCTGGATGACCTGACGCTGGAGGGCATCATCGGATTCGGTGACATGGCCGCCACCGCCGACA
CGTCCATGATGTGCGGCCGCGGCATCCACGACTCGAAGTTTCCCGGCCCAACGACCGCCGCC
35 GTCTGGGCTTCGTCGAAGCCCAAGGCGGCGGGACGATCCTGTTGGCCCGCGGGGACCTGGCG
CTGCGGATGGGGCTGCCGGTGCTGGCGGTGGTGGCGTTCGCGCAGTCGTTCCGGCGACGGCGT
GCACACCTCGATCCCGGCCCGGGCCTGGGCGCGCTGGGGCGGGCCCGCGCGGCAAGGAT

TCACCGCTGGCGCGGGCGCTGGCCAAGCTGGGCGTGGCCGCCGACGACGTGGCGGTTCATCTC
CAAGCACGACACCTCGACGCTGGCCAACGATCCCAACGAGACCGAGTTGCATGAACGGCTCGC
CGACGCCCTGGGCCGTTCCGAGGGCGCCCCGCTGTTCTGTGGTGTGCGAGAAGAGCCTGACCG
GCCACGCCAAGGGCGGGCGGGCGGTCTTCCAGATGATGGGGCTCTGCCAGATATTGCGGGAT
5 GGGGTGATCCCAACCGCAGCCTCGACTGCGTCGACGACGAGCTGGCCGGCTCCGCGCA
TTTCGTGTGGGTGCGTGACACGTTGCGGCTCGGCGGCAAGTCCCCTCAAGGCCGGCATGCT
GACCAGCCTCGGGTTTCGGCCATGTGTGGGCTCGGCGGCAAGTCCCCTCAAGGCCGGCATGCT
CGCCTCGCTGGATCCCGCACAGCGCGGGACTACCAGCGGCGTGCCGACGCCCGCCTGCTGG
CCGGTCAGCGCCGGCTGGCCTCGGCGATTGCCGGTGGTGGCGCGATGTACCAGCGGCCCGGT
10 GACCGTCGCTTCGACCACCACGCGCCCCGAGCGGCCGAGGAGGCGTCGATGCTGCTGAATCC
GGCGGCCCGGCTGGGTGACGGCGAGGCGTATATCGGCTGA

>Rv2555c alaS alanyl-tRNA synthase TB.seq 2873772:2876483 MW:97326

>emb|AL123456|MTBH37RV:c2876483-2873769, alaS SEQ ID NO:101

15 GTGCAGACACACGAGATCAGGAAGCGGTTCTCGATCATTTCTGTGAAGCGGGCCACACCGAG
GTGCCCAGCGCCTCGGTGATCCTCGACGACCCCAACCTGTTGTTCTGTCAACGCCGGGATGGTC
CAGTTCGTGCCTTTCTTCTTGGGACAGCGCACGCCGCCGTACCCGACGGCCACCAGCATCCAG
AAGTGCATCCGTACCCCCGATATCGACGAGGTGGGCATAACCACCCGGCACAACACGTTTTTTC
AGATGGCCGGCAATTTTCAGCTTCGGCGACTATTTCAAACGCGGGGCCATTGAACTGGCCTGGG
20 CACTGCTGACCAACAGCCTCGCCGCCGGCGGCTACGGCCTGGACCCGAAAGAATCTGGACG
ACAGTCTATTTTCGACGACGACGAAGCTGTCCGGCTATGGCAGGAGGTTGCCGGGCTGCCGGCG
GAGCGAATCCAGCGCCGCGGCATGGCCGACAACTACTGGTCGATGGGCATTCCCGGACCGTG
CGGGCCGTCATCGGAGATCTATTACGACCGCGGACCCGAATTCGGTCCCGCAGGCGGTCCCAT
CGTCAGCGAAGACCGCTACCTCGAGGTCTGGAACCTGGTGTTCATGCAGAACGAGCGCGGAGA
25 GGGAACCACCAAGGAGGACTACCAGATCCTCGGGCCGCTGCCCCGCAAGAACATCGACACCG
GCATGGGCGTCGAGCGGATCGCGCTGGTGTGCAAGACGTGCACAACGTCTACGAGACCGAC
CTGCTCAGGCCGGTCATCGATACCGTGGCCAGGGTCGCCGCGCGTGCCTACGACGTGCGCAA
CCACGAAGACGACGTGCGGTACCGCATCATCGCAGACCACAGCCGCACCGCCGCGATCCTGAT
CGGTGACGGCGTCAGCCCCGGCAACGACGGTTCGCGGTTATGTGCTGCGCCGGCTGCTGCGTC
30 GGGTGATCCGCTCCGCCAAGCTGCTGGGCATCGACGCTGCGATCGTTGGCGACCTGATGGCCA
CGGTGCGCAACGCGATGGGCCCCGTCATATCCCGAACTCGTCGCCGACTTCGAGCGGATCAGCC
GGATCGCGGTGCGCGAGGAGACGGCGTTCAACCGCACGCTGGCGTCGGGTTCAGGCTGTTTC
GAGGAGGTGGCTAGCTCCACCAAGAAATCCGGAGCCACCGTGCTGTCCGGATCGGACGCTTTC
ACGTTGCATGACACCTACGGGTTCCCGATCGAGCTCACGCTGGAGATGGCGGCCGAAACCGGT
35 CTGCAGGTAGACGAAATCGGGTTCGCTGAGCTGATGGCCGAGCAGCGCCGCCGTGCCAAGGC
CGACGCCGCCGCGCGCAACACGCGCATGCTGACCTGAGCGCCTACCGCGAGCTGGTTGACG
CCGGCGCCACCGAGTTCACCGGATTCGACGAGTTGCGTTCCAGGCGCGGATTCTGGGCATCT

TCGTCGACGGTAAGCGGGTTCGGTGGTGGCGCACGGTGTAGCCGGCGGAGCCGGGGAAGG
GCAGCGTGTGAACTTGTCTTAGATCGCACCCCGCTCTACGCCGAATCGGGTGGGCAGATCGC
CGATGAGGGCACCATCAGCGGAACCGGTTCAGCGAAGCTGCCCGGGCCGCGGTACCGACG
TGCAGAAGATCGCCAAAACGCTTTGGGTGCACCGAGTCAACGTGGAATCCGGGGAATTCTGTCG
5 AGGGTGACACCGTAATCGCGGCGGTGGATCCCGGGTGGCGCCGGGGTGCCACGCAGGGCCA
CTCGGGCACCCACATGGTGCATGCCGCGCTGCGACAAGTGCTGGGGCCCAACGCGGTTCAGG
CGGGATCGCTGAACCGGCCGGGATATTTGCGCTTCGACTTTAACTGGCAGGGTCCGTTGACCG
ACGACCAGCGCACCCAGGTGGAAGAGGTCAACAACGAGGCCGTGCAAGCGGACTTCGAGGTG
CGCACGTTACCGAACAGCTCGACAAGGCCAAGGCGATGGGTGCCATCGCGCTGTTCCGGCGAG
10 AGCTACCCCGACGAAGTGCGGGTGGTGGAGATGGGTGGACCGTTCTCGCTGGAGCTATGTGGC
GGCACCCATGTGAGCAACACGGCGCAGATCGGTCCCGTGACGATCCTGGGCGAGTCGTGATC
GGCTCCGGGGTGCGCCGGGTGGAGGCCTACGTGGGGTTGGATTCTTTCTGTCACCTGGCCAA
GGAGCGTGCGTTGATGGCCGGGTGGCCCTCGTCACTGAAGGTGCCGTCCGAAGAGGTACCGG
CCCGGGTGGCCAATCTAGTGGAGCGCCTGCGGGCCGCCGAGAAGGAACTCGAACGTGTCCGG
15 ATGGCCAGCGCCCGGGCAGCCGCCACCAATGCCGCCGCCGGGGCTCAGCGGATCGGTAACGT
CCGTTTGGTGGCGCAGCGAATGTCCGGCGGGATGACCGCGGCAGACCTGCGGTGCGTTGATCG
GCGACATCCGCGGCAAGCTGGGTAGCGAGCCGGCGGTGGTGGCGCTGATTGCCGAGGGCGAA
AGCCAAACTGTGCCGTATGCGGTGCGGGCCAATCCCGCTGCCCAGGACCTCGGAATCCGTGCC
AACGACCTGGTCAAACAATTGCGGTGGCGGTGGAAGGCCGCGGTGGCGGTAAGGCGGACCT
20 GGCGCAGGGCTCGGGAAAGAATCCGACCGGTATCGACGCCGCGCTCGACGCGGTCCGCTCCG
AGATCGCCGTGATAGCGCGGGTCCGTTGA

>Rv2580c hisS histidyl-tRNA synthase TB.seq 2904822:2906090 MW:45118

>emb|AL123456|MTBH37RV:c2906090-2904819, hisS SEQ ID NO:102

25 GTGACGGAATTCTCGTCATTTTCGGCCCCCAAGGGGTACCGGACTACGTCCCGCCCGACTCG
GCGCAGTTCGTGCGGGTGCAGCGACGGGTGCTCGCGGCGGCCCGTCAAGCCGGCTATAGCCA
CATCGAGCTGCCCATCTTCGAGGACACCGCCCTGTTCCGCCGGGGCGTGGGTGAATCCACCGA
CGTGGTGTCCAAGGAGATGTATACGTTCCCGACCGTGGCGACCGCTCGGTGACGCTGCGGCC
CGAGGGCACCGCCGGGTGGTGCCTGCGGTGATCGAACACGGGCTGGATCGCGGCGCGCTG
30 CCGGTGAAGTTGTGTTATGCGGGCCCGTTCCTCGCTACGAGCGTCCGCAGGCCGGCCGGTAT
CGCCAGTTACAGCAAGTCGGGGTGGAGGCGATCGGCGTCGACGACCCGGCGTTGGACGCCGA
GGTGATCGCCATTGCCGACGCCGGGTTCGCTCGTTGGGTCTCGACGGGTTCGGGCTGGAAAT
CACCTCCCTGGGAGACGAGAGTTGCCGTCCGAGTACCGGGAACTGTTGCAGGAGTCTTGTTT
GGACTCGATCTCGACGAGGACACCCGAGGCGCGCAGGGATCAATCCGCTGCGGGTGTCTGA
35 CGACAAGCGACCCGAATTGCGTGCGATGACGGCGTGGCGCCGGTGTGCTGGATCATCTGTC
TGATGTCGCCAAGCAGCATTTCGACACCGTGCTCGCCCATCTGGACGCGCTTGAGTGCCCTAT
GTCATCAACCCGCGCATGGTGCAGGCGCCTGGACTACTACACCAAGACCGCCTTCGAGTTCGTC

CATGACGGGCTTGGTGCGCAATCGGGGATCGGCGGCGGGGGGCGCTACGACGGCCTGATGCA
CCAGCTTGGCGGGCAGGACTTGTGCGGCATCGGGTTCGGGCTGGGCGTGGACCGGACCGTGC
TGGCGCTGCGGGCCGAGGGCAAGACGGCGGGGGACAGCGCCCGGTGCGACGTGTTGCGCGT
GCCGCTTGGCGAGGCGGCCAAGCTCAGGCTGGCGGTGCTGGCTGGACGACTGCGCGCGGGCC
5 GGGGTGCGGGTTGACCTTGCTATGGTGATCGCGGGCTCAAAGGCGCGATGCGCGCGGGCCG
TCGTTCCGGCGCCCGTGTTCGTTGGTAGCGGGCGACCGCGACATCGAGGCCGGGACGGTCG
CAGTGAAGGACTTGACGACGGGTGAGCAAGTTTCGGTCTCGATGGATTGCGTTGTGGCCGAAG
TAATTCGCGGCTGGCTGGGTAG

10 >Rv2614c thrS threonyl-tRNA synthase TB.seq 2941190:2943265 MW:77123
>emb|AL123456|MTBH37RV:c2943265-2941187, thrS SEQ ID NO:103
ATGAGCGCCCCCGCACAAACCCGCCCCGGGAGTCGATGGCGGCGACCCGTGCAAGCCCGAAT
TCGGGTTCCCTGCCGGGACCACCGCGGCCACCGCCGTGCGCGAAGCGGGTTTACCGCGGGCGCG
GTACGCCCCGATGCGATCGTCGTCGTGCGCGACGCCGACGGCAACCTGCGCGACCTGAGCTGG
15 GTGCCCCGACGTCGACACCGATATCACGCCGGTGGCCGCCAACACCGACGACGGTCGCAGCGT
GATCCGCCATTGACCGCGCACGTGTTGGCCCAAGCCGTCCAAGAGCTGTTTCCCGAGGCCAA
GCTCGGCATCGGACCACCCATCACCGACGGCTTCTACTACGACTTCGACGTGCCCGAGCCGTT
CACGCCCCGAGGACTTGCGGGCGCTGGAAAAGCGGATGCGCCAGATCGTCAAGGAAGGCCAGC
TGTTGACCGGGCGGGTCTACGAATCCACCGAACAGGCCCGCGCCGAGCTGGCCAACGAGCCC
20 TACAAGCTGGAACTCGTCGACGACAAATCGGGTGACGCCGAGATCATGGAGGTGCGCGGTGAC
GAGCTCACCGCCTACGACAACCTCAACCCCCGCACCCGCGAGCGCGTCTGGGGCGACCTGTG
CCGCGGACCGCACATCCCGACCACCAACACATCCCGGCGTTCAAGCTCACCGCAGCTCGGC
CGCCTACTGGCGGGGCGATCAGAAAAACGCCAGCCTGCAACGGATCTACGGCACCGCGTGGG
AATCCCAGGAGGCGCTCGACAGGCACCTGGAGTTCATCGAAGAGGCGCAGCGCCGCGACCAC
25 CGCAAGCTGGGTGTCGAGCTGGACCTGTTAGCTTCCCCGACGAAATCGGTTCCGGCCTAGCG
GTTTTCCACCCCAAGGGCGGCATCGTGCGTTCGCGAACTGGAGGACTACTCGCGGCGCAAGCAC
ACCGAGGCGGGCTACCAAGTTCGTCAACAGCCCGCACATCACCAAGGCCAGTTGTTCCACACC
TCGGGACATCTGGAAGTGGTACGCCGACGGCATGTTCCCCCGATGCACATCGACGCGGAGTAC
AACGCCGACGGCTCGCTGCGCAAACCCGCGCAGGACTACTACCTCAAGCCGATGAACTGCCCG
30 ATGCACTGCCTGATCTTCCGCGCGCGCGGGCGATCCTATCGGGAAGTCCGTTGCGGCTCTTC
GAGTTCGGCACGGTGTATCGCTACGAGAAGTCCGGTGTGTCACGGGTGACCCGGGTGCGT
GGGCTGACCATGGACGACGCGCACATCTTCTGCACCCGCGACCAGATGCGCGACGAGCTGCG
GTCGCTGCTGCGGTTTGTGCTCGACCTGCTCGCCGACTACGGCCTCACCGACTTCTACCTCGAA
CTGTCCACCAAGGACCCGAGAAAGTTCGTGCGCGCCGAGGAGGTCTGGGAGGAAGCCACCAC
35 CGTGCTGGCCGAGGTGGGCGCCGAATCCGGGCTGGAGCTGGTGCCCGATCCAGGCGGCGCG
GCGTTCTACGGGCCCAAGATTTCAGTGCAGGTCAAAGACGCGCTGGGCGCGACCTGGCAGATG
TCGACCATCCAGCTGGACTTCAACTTTCGGGAACGTTTCGGCCTGGAGTACACCGCCGCGGACG

GAACCCGCCACCGCCCCGGTGATGATCCACCGCGCGCTATTTGGGTCGATCGAGCGGTTCTTCG
GCATTCTCACCGAGCACTACGCGGGGGCGTTCCCGGCCTGGTTGGCGCCCGTGACGGTGTC
GGCATCCCGGTCGCCGATGAGCACGTGCGCTATCTGGAAGAGGTTGCCACGCAACTGAAGTCG
CACGGGGTGCGGGCCGAGGTGGACGCCAGCGACGATCGGATGGCCAAGAAGATCGTGACCA
5 CACCAACCACAAGGTGCCGTTTCATGGTGTGGCGGGTGATCGTGACGTGCGCCGCCGGCGCGGT
GAGTTTCCGGTTCGGTGACCGCACCCAAATCAACGGTGTGGCCCGTGACGATGCGGTGGCGGC
CATTGTGCGCTGGATCGCTGACCGCGAAAATGCGGTTCTACAGCGGAACTGGTGAAAGTGGC
CGGTCTGAGTGA

10 >Rv2697c dut deoxyuridine triphosphatase TB.seq 3013683:3014144 MW:15772
>emb|AL123456|MTBH37RV:c3014144-3013680, dut SEQ ID NO:104
GTGTCGACCACTCTGGCGATCGTCCGCCTCGACCCCGGGCTCCCGCTGCCAGCCGCGCTCAC
GACGGCGACGCCGGCGTTGATCTCTACAGCGCCGAAGACGTGAGCTGGCACCTGGGCGCCG
CGCCCTGGTACGGACGGGTGTTGCGGTGCGCGTCCCGTTCGGCATGGTCGGGCTGGTCCATC
15 CGCGCTCCGGGTTGGCCACGCGGGTGGGGCTTTGATCGTCAACAGTCCGGGCACCATCGAC
GCGGGTTATCGTGGGGAGATCAAGGTGGCCCTGATCAACTTGGACCCAGCCGCGCCCATCGTG
GTACATCGCGGTGACCGAATCGCCAGTTGCTAGTGCAACGGGTGAGTTGGTCGAGCTGGTC
GAGGTCTCGTCGTTGACGAGGCCGGGCTGGCCTCGACATCCCGCGGCGACGGTGGCCACGG
TTCCTCCGGCGGACATGCGAGTTTGTGA

20 >Rv2782c pepR protease/peptidase, M16 family (insulinase) TB.seq 3089045:3090358 MW:47074
>emb|AL123456|MTBH37RV:c3090358-3089042, pepR SEQ ID NO:105
ATGCCGCGACGGTCACCAGCTGACCCCGCGGCGGCGCTGGCGCCGCGGCGCACCAACCTGC
CGGGCGGGCTGCGAGTGTCACCGAATTCCTGCCCGCGGTGCACTCCGCGTCGGTCGGGGTG
25 TGGGTGCGCGTCGGATCGCGCGACGAAGGCGCCACGGTGGCCGGGGCGGCGCACTTCCTTGA
GCATTTGCTGTTCAAGTCGACGCCACCCGCTCTGCCGTGGACATTGCGCAGGCGATGGACGC
GGTGGGCGGGGAAGTGAACGCATTACCGCCAAGGAGCACACCTGCTACTACGCCACGTGCT
CGGCAGCGACTTGCCGTTGGCCGTGACCTGGTCGCCGATGTGGTGCTCAACGGCCGCTGTGC
CGCCGACGATGTCGAGGTGGAACGTGACGTGTCCTCGAGGAGATCGCGATGCGCGACGACG
30 ACCCGGAGGACGCCCTTGGCGGACATGTTCTGGCGGCGTTGTTGCGCGACCAACCGGTGCGTC
GCCCGGTGATCGGCAGCGCGCAATCCGTGTGGTGATGACGCGGGCTCAACTGCAATCGTTTC
ACCTGCGGCGCTATACCCCGGAGCGGATGGTCGTGCGGGCCGCGGCAATGTGGATCACGAC
GGGCTGGTTGCGTTGGTCCGCGAGCACTTCGGGTCCCGGTTGGTCCGGGGGAGACGGCCAGT
TGCGCCGCGCAAGGGTACCGGCCGGGTCAACGGCAGCCCCCGGTTGACACTGGTTAGCCGCG
35 ACGCCGAACAGACGCATGTGTGCTGGGCATCCGCACACCCGGGCGCGGCTGGGAGCATCGT
TGGGCACTGTCGGTGCTGCACACCGCGCTGGGCGGTGGCTTGAGTTCCCGGCTGTTCCAGGAG
GTCCGCGAGACCCGCGGGCTGGCCTACTCGGTCTACTCCGCGCTGGATCTCTTCGCCGACAGC

GGCGCGCTTTCCGGTGTACGCGGCCTGCCTGCCCCAACGCTTCGCCGACGTGATGCGGGTGAC
CGCCGATGTGCTGGAAAGCGTGGCACGCGACGGCATCACCGAGGCGGAATGCGGCATCGCCA
AGGGATCGCTGCGGGGTGGGCTGGTGCTAGGGCTGGAGGATTCCAGCTCCCGGATGAGCCGG
CTCGGCCGCGAGCGAGTTGAACTACGGCAAGCACCGCAGCATCGAACACACCTTGCGGCAAATC
5 GAGCAGGTACCCGTGGAGGAGGTCAACGCGGTGGCCCCCACCTGCTGAGCAGGCGCTACGG
TGCTGCCGTTCTTGCCCCACACGGATCGAAACGATCACTGCCGCAACAACTTCGAGCGATGGTA
GGGTAG

>Rv2783c gpsi pppGpp synthase and polyribonucleotide phosphorylase TB.seq

3090339:3092594 MW:79736 >emb|AL123456|MTBH37RV:c3092594-3090336, gpsi

SEQ ID NO:106

ATGTCTGCCGCTGAAATTGACGAAGGCGTGTTGAGACGACCGCCACCATCGACAACGGGAGC
TTTGGCACCCGGACCATCCGCTTCGAGACCGGCCGATTGGCCTTGACAGGCCGCCGGCGCGGT
GGTCGCCTACCTCGACGACGACAACATGCTGCTGTCGGCGACCAACCGCCAGCAAGAACCCCAA
15 AGAACACTTCGACTTCTTCCCCCTCACGGTCGACGTCGAGGAGCGCATGTATGCGGCCGGCCG
CATCCCCGGTTCGTTCTTCCGTGCGGAGGGCCGACCTCCACCGACGCGATCCTGACCTGCCG
GCTCATCGACCGCCCCGTGCGCCCCGTGTTTTGTCGACGGGCTGCGCAACGAGATCCAAATCGT
GGTGACGATTCTCAGCCTGGATCCGGGCGATCTCTACGACGTATTGGCGATCAACGCGGCGTC
GGCGTCCACCCAGCTGGGCGGTCTGCCGTTCTCCGGGCCCATCGGCGGTGTGCGGGTGGCGC
20 TCATCGACGGCACCTGGGTGCGGTTCCCCACCGTCGACCAGATCGAGCGCGCCGTGTTGACA
TGGTCGTGGCCGGCCGGATCGTCGAGGGTGATGTTGCCATCATGATGGTCAAGCCGAGGCCA
CCGAAAACGTCGTCGAGCTCGTCGAAGGTGGTGCCCAAGCGCCGACGGAAGCGTGTTGGCC
GCGGGCCTGGAGGCGGCCAAGCCGTTATCGCCGCGCTGTGCACCGCGCAGCAGGAGCTTGC
CGATGCCGCTGGAAGTCGGGCAAACCGACCGTCGACTTCCCGGTGTTCCCTGACTACGGCGA
25 AGACGTGTACTACTCGGTGTCTCGGTGGCCACCGACGAGTTGGCCGCCGCGTTGACCATCGG
CGGTAAGCCGAGCGCGACCAAGCGCATCGACGAAATCAAGACCCAGGTTGTGACGCGGCTCGC
CGACACCTACGAGGGTCGCGAAAAGGAGGTGCGCGCCGCGTTGCGTGCCCTGACCAAAAAGCT
GGTTCGGCAGCGCATCCTCACCGACCATTTCCGTATCGACGGCCGCGGCATCACCGACATTCG
CGCATTGTGCGCCGAGGTGGCCGTGGTTCCGCGCGCGCACGGCAGCGCGCTGTTGAAACGCG
30 GCGAAACCCAGATCCTGGGTGTGACCACACTCGACATGATCAAGATGGCCCAGCAGATCGACT
CGTTGGGGCCGGAGACATCGAAGCGGTACATGCACCACTACAACTTCCCGCCGTTCTCCACCG
GCGAGACCGGTGCGGTGCGTTCCGCCAAGCGGCGTGAGATCGGGCACGGCGCACTGGCCGA
GCGGGCCCTGGTGCCGGTGTGCGGAGCGTCGAGGAATCCCGTATGCCATTGCGCAGGTGTC
GGAGGCTCTGGGCTCCAACGGGTGACCTCGATGGGGTGGTGTGCGCGTCGACGCTGGCGC
35 TGCTCAACGCCGGGGTGCCGCTCAAGGCGCCGGTGGCCGGCATCGCGATGGGCCCTGGTCTCC
GACGACATTCAAGTAGAAGGGGCGGTGACGCGGCTGTGGAGCGTCGCTTCGTACCCTCACC
GACATCCTCGGCGCCGAAGACGCGTTCGGTGACATGGACTTCAAGGTGCGCCGGGACCAAGGAC

TTCGTCACCGCGCTGCAGCTGGACACCAAGCTCGACGGGATCCCTTCGCAGGTGCTTGCCGGA
GCACTCGAGCAGGCCAAGGACGCCCGCCTCAGGATCTTGAGGTGATGGCTGAGGCCATCGAT
AGACCCGACGAAATGAGTCCCTACGCCCCGCGGGTGACCACCATCAAGGTTCCGGTGGACAAG
ATCGGGGAGGTCATCGGACCCAAGGGCAAGGTCATCAACGCCATCACCGAGGAGACCGGGCGC
5 GCAGATCTCCATCGAAGACGACGGCACCGTGTTCGTCGGCGCCACCGACGGGCCATCGGCACA
GGCCGCGATCGACAAGATCAACGCCATCGCCAACCCGACAGCTGCCGACGGTGGGCGAACGGT
TCCTCGGAACCGTGGTCAAGACCACCGATTTCGGTGCCTTTGTATCGTTGCTGCCTGGCCGCGA
CGGTCTGGTGACATTTCCAAACTCGGCAAGGGCAAGCGCATCGCGAAGGTGAGGACGTTGT
CAATGTCGGTGACAAGCTGCGGGTGGAGATCGCCGACATCGACAAACGGGGCAAGATCTCCCT
10 GATCCTGGTCGCGGACGAGGACAGCACCGCCCGCGCTACCGATGCCGCGACGGTCACCAAGT
GA

>Rv2793c truB tRNA pseudouridine 55 synthase TB.seq 3102364:3103257 MW:31821

>emb|AL123456|MTBH37RV:c3103257-3102361, truB SEQ ID NO:107

15 ATGAGCGCAACCGGCCCGGAATCGTGGTTATCGACAAGCCCGCGGGAATGACCAGCCATGAC
GTGGTGGGGCGGTGCCGCCGCATCTTCGCCACCCGGCGGGTCGGCCACGCGGGCACCCCTGG
ACCCGATGGCCACCGGGGTGTTGGTGATCGGCATCGAACGCGCCACCAAGATCCTCGGTCTGC
TGACGGCGGCCCCCAAGTCGTATGCCGCCACCATCCGCTTGGGTCAGACCACTTCCACCGAGG
ACGCCGAAGGTCAAGTGCTGCAGTCGGTTCGGGCTAAGCACCTGACCATCGAGGCGATCGACG
20 CCGCGATGGAGCGGCTGCGCGGTGAGATCCGGCAGGTGCCGTGCTCGGTCAGCGCGATCAAG
GTCGGTGGCCGACGCGCCTATCGGTTGGCCCCGCCAGGGCGCTCCGTGCAATTGGAAGCCCG
GCCGATCCGCATCGACCGGTTGAGCTGCTGGCCGCACGCCGGCGCGACCAAGCTCATCGATAT
CGATGTGGAGATCGACTGCTCCTCGGGAACCTACATCCGCGCGTTGGCACGCGACCTCGGCCA
CGCGCTTGGGGTGGGAGGCCATGTGACGGCGTTGCGGCGACCCGCGTCGGCCGCTTCGAGC
25 TGGACCAGGCGAGATCGCTCGACGATCTCGCGGAGCGCCCCGCGCTGAGCCTGAGCCTCGAT
GAGGCCTGCCTGCTGATGTTTGCGCGCCGCGACCTGACCGCCGCGGAGGCCAGCGCGGCCGC
CAACGGCCGGTCCCTGCCGGCGGTGCGGTATCGACGGCGGTACGCGGCCTGTGACGCCGACG
GCCGGGTTATCGCGCTGCTGCGTGACGAGGGTTCGCGGACCAGGTCGGTGGCGGTGCTCCGC
CCGGCGACGATGCACCCCGGGTAG

30 >Rv2797c - TB.seq 3105619:3107304 MW:58761 >emb|AL123456|MTBH37RV:c3107304-3105616,
Rv2797c SEQ ID NO:108

GTGCCACTGACCGTGGCCGATATCGATCGGTGGAACGCGCAAGCGGTCCGGGAGGTGTTTCAC
GCGGCCAGTGCCCGAGCGGAGGTGACGTTGAGGCGTCGCGTCAGTTGGCCGCGCTGTCGAT
35 TTTTGCGAACTCGGGTGGCAAGACCGCTGAGGCGGCGGCACACCACAACGCGGGCATTGCC
GAGACCTCGACGCCACGGCAACGAGGCGTTGGCGGTTGCCCGGGCGGCCGACAGGGCCGC
CGACGGGATTGTGAAGGTTCACTCCGAGCTGCCCGCACTACGCCATGCCGCCGCGGCCGCCG

AGCTGACGATCGATGCGCTGATCAACCGGGTGGTGCCGATCCCCGGGCTGCGATCCACCGAG
GCGCAGTGGGCGCGGACGCTGGCCAAGCAAACGGAGCTGCAGGCGGAGCTGGATGCGATTAT
GGCCGAGGCCAATGCCGTCGACGAGGAGCTGGCCTCAGCGGTCAATATGGCCGACGGTGACG
CGCCCATCCCGGCCGATTCCGGCCCGCCGGTCCGTCCCGAGGGGGCTGACCCCGACCCAGCTC
5 GCCAGCGATGCCAACGAGGAGCGGCTGCGCGAGGAGCGCGCCCGCCTGCAGGCCACCTCG
AGCGGTTACAGGCGGAGTATGACCAACTGAGTGTGCGGGCCGCCCCGTGACTACCACAACGGCA
TCCTCGACGGTGACGCGGTGGGCCGACTGGCAGCGCTTACCGACGAGCTGAGCGCCGCCAGG
GGCCGGCTGGGTGAGCTCGATGCCGTCGACGAGGCGTTGAGCCGAGCACCCGAGACCTACCT
GACCCAGCTGCAGATTCCCGAGGACCCAAATCAGCAGGTGCTGGCGGCCGTGGCCGTCGGTAA
10 TCCCGACACCGCCGCCAATGTGTGCGGTGACGGTTCCCGGCGTCGGGTCCACCACCCGGGGCG
CCCTGCCCGGCATGGTGACCGAAGCCCGCGACCTGCGGTGCGGAGGTAATCCGGCAACTCAATG
CTGCCGGCAAGCCCGCATCGGTTGCCACCATCGCCTGGATGGGCTACCACCCGCCCCCGAACC
CACTCGACACCGGCAGTGCGGGCGATCTGTGGCAGACCATGACCGATGGGCAGGCACACGCG
GGCGCGGCCGATCTGTGCGGGTATTTGCAGCAGGTGCGCGCCAATAACCCAGTGGCCACCTG
15 ACCGTGTTGGGGCACTCGTATGGGTGCTGACGCGCTCGCTGGCGTTGCAGGACCTCGATGCC
CAGAGCGCCCATCCCGTCAACGACGTGCTGTTTTACGGCTCACCCGGCTTGGAGCTGTACAGC
CCGGCGCAGCTCGGGCTCGATCACGGGCACGCTTATGTCATGCAGGCCCCCACGACCTCATC
ACCAATCTGGTGGCGCCGTTGGCGCCGCTGCACGGATGGGGCCTGGACCCCTATCTGACCCCC
GGGTTACGGAGCTGTGCTCACAGGCGGGTTTTGATCCGGGCGGGATCTGGCGTGACGGAGT
20 GTATGCCACGGGGACTACCCGCGGTCTTCTCGATGCCGCGGCCAGCCGCAGCTGCGGA
TGTCCGGCTATAACCTGGCGGCGATCGCCCGGGGCTGCCCGACAACACGGTGGGCCCGCCG
CTGCTTCCGCCAATTCTGGGTGGCGGCATGCCGBCAGCGCCCGGCCAGCACTGAGAGGGGG
ACGTTGA

25 >Rv2864c ponA2 TB.seq 3175454:3177262 MW:63015 >emb|AL123456|MTBH37RV:c3177262-
3175451, Rv2864c SEQ ID NO:109
ATGGTAACTAAAACAACATTAGCCTCAGCCACCTCAGGTTTGCTGCTGCTTGCGGTGCTCGCCAT
GTCGGGCTGCACCCCGCGTCCCCAAGGGCCCGGTCCGGCGGCCGAAAAGTTCTTCGCCGCGC
TGGCCATCGGTGACACCGCCTCCGCCGCCAGCTCAGCGACAACCCCAACGAGGCGCGCGAA
30 GCGCTGAACGCGGCCTGGGCGGGGCTGCAGGCCGCCACCTGGATGCGCAGGTTCTCAGCGC
CAAGTACGCCGAGGACACCGGTACGGTCGCTTATCGCTTCAGCTGGCATCTGCCAAGGACCG
AATCTGGACCTATGACGGCCAGCTGAAGATGGCCCGCGACGAAGGGCGTTGGCACGTTTCGCTG
GACCACCAGCGGGTTGCATCCCAAGCTAGGCGAACATCAAACGTTTCGCGCTACGAGCCGACCC
GCCGCGGCGCGCCTCGGTGAACGAAGTCGGCGGCACCGATGTGCTGGTGCCGGGCTATCTGT
35 ATCACTACTCGCTGGACGCCGCGCAGGCCGCGCGAGCTCTTCGGCACGGCACACGCGGTG
GTGGGCGCGCTGCACCCCTTCGACGACACGCTCAATGATCCGCAGCTGCTGGCCGAACAGGCC
AGCTCGTCGACCCAGCCGTTGGACCTGGTCACGTTGCACGCCGACGACAGCAACCGGGTGGC

CGCGGCGATCGGGCAGCTGCCTGGCGTGGTGATCACACCGCAGGCCGAGCTGCTCCCGACCG
ACAAGCACTTCGCGCCGGCGGTCTCAACGATGTCAAGAAGGCCGTCGTCGATGAACTCGACG
GCAAGGCGGGTTGGCGGGTGGTGAGCGTCAACCAAAATGGCGTCGACGTCTCGGTGCTGCAC
GAGGTCGCCCCATCACCTGCGTCGTCGGTTTCGATCACGTTGGATCGGGTCGTGCAAAACGCC
5 GCGCAACACGCGGTGAACACCCGGGGCGGCAAGGCGATGATCGTCGTGATCAAGCCGTCGAC
CGGCGAGATCCTGGCGATCGCGCAGAACGCCGGGGCCGATGCGGACGGTCCGGTCGCGACCA
CCGGTCTATATCCACCCGGGTGACATTCAGATGATCACCGCCGGTGCGGCCGTCGAGCGTG
ACCTGGCTACCCCTGAGACGCTGCTGGGTTGCCCGGGGAGATCGACATCGGGCATCGCACCA
TTCCCAACTACGGTGGCTTTGATCTGGGCGTGGTGCCGATGTCACGCGCGTTTGCCAGTTCCTG
10 CAACACCACCTTCGCCGAGCTGAGCAGCAGGCTGCCTCCCCGCGGTCTGACTCAGGCGGCCCC
GGCGGTACGGGATCGGGCTTGACTACCAGGTGGACGGCATCACCACGGTGACCGGTTCCGGTG
CCGCCGACGGTGGACCTGGCCGAACGCACCGAGGACGGTTTCGGCCAGGGCAAGGTGCTGGC
CAGCCCGTTCCGGCATGGCCTTGGTGGCGGCGACGGTAGCCGCCGGGAAGACCCCGGTTCCAC
AGCTGATCGCCGGCCGGCCGACGGCCGTGCAAGGCGATGCCACACCGATCAGCCAGAAGATG
15 ATCGACGCGCTGCGGCCCATGATGCGGTTGGTGGTGACCAATGGCACCGCCAAGGAGATCGCT
GGCTGTGGCGAGGTGTTCCGTAAGACCGGCGAAGCCGAATTCCCGGGCGGATCGCATTCCTG
GTTCCGCCGGTACCGTGGCGATCTGGCATTGCGTCGCTGATCGTCGGGGGCGGTAGCTCGGA
ATACGCGGTGCGGATGACCAAGGTGATGTTCAATCGCTGCCGCCGGGGTACCTGGCGTAG

20 >Rv2868c gcpE TB.seq 3179368:3180528 MW:40451 >emb|AL123456|MTBH37RV:c3180528-
3179365, gcpE SEQ ID NO:110
GTGACTGTAGGCTTGGGCATGCCGCAGCCCCCGGCACCCACGCTCGCTCCCCGGCGCGCCAC
CCGTCAGCTGATGGTCGGCAACGTCGGCGTGGGCAGTGACCATCCGGTCTCGGTGCAATCGAT
GTGCACCACCAAAACCCACGACGTCAACTCGACATTGCAACAAATCGCCGAGCTGACCGCGGC
25 CGGATGCGACATCGTGCGGGTGGCCTGCCCGCGCCAGGAGGACGCCGACGCGCTGGCCGAG
ATCGCCCCGGCACAGCCAGATCCCGGTAGTCGCGGACATACATTTCCAGCCGCGCTACATATTG
CCGCCATCGACGCTGGATGTGCCGCGGTGCGGGTCAACCCGGGCAACATCAAGGAGTTTGACG
GCCGGGTGGGTGAGGTGCCAAGGCGGCGGGTGGCGCCGGGATCCCGATCCGAATCGGTGT
CAACGCCGGTTCCGCTGGACAAACGGTTCATGGAGAAGTATGGCAAAGCCACGCCCGAGGCGCT
30 GGTGAGTCGGCGCTGTGGGAGGCTTCGCTTTTCGAGGAGCATGGCTTCGGTGACATCAAGAT
CAGCGTCAAGCACAACGACCCGGTGGTGTGGTTCGCCGCTACGAGCTGCTTGCTGCACGGTG
CGACTACCCACTGCACCTCGGTGTACCGAGGCGCGCCCTGCTTTCCAGGGCACCATCAAGTC
CGCGGTTGCCCTTCGGCGCGTTGCTGTGCGGGGCATAGGCGACACCATCCGGGTGTGCTTGTC
GGCCCCGCCGGTCGAGGAAGTCAAGGTGGGCAATCAGGTTCTCGAGTCGTTGAACCTGCGGGCC
35 GCGTTCGCTCGAGATCGTGTCTTGCCCGTCGTGCGGTGCGCGCAAGTCGACGTCTACACCT
GGCCAACGAGGTAACCGCCGGCCTGGATGGTCTCGATGTGCCGTTGCGGGTGGCGTGATGG
GGTGTGTCGTCAATGGTCCGGGTGAAGCACGTGAGGCCGACCTGGGCGTGGCGTCCGGCAAC

GGCAAAGGTCAGATCTTTGTACGGGGCGAAGTGATCAAGACCGTGCCCGAAGCACAGATCGTC
GAGACGCTGATCGAGGAGGCGATGCGGCTGGCCGCCGAAATGGGCGAGCAAGATCCGGGCGC
GACACCGAGCGGTTGCGCTATTGTGACCGTAAGCTGA

5 >Rv2869c - TB.seq 3180548:3181759 MW:42835 >emb|AL123456|MTBH37RV:c3181759-3180545,
Rv2869c SEQ ID NO:111
ATGATGTTTGTACCGGCATTGTGCTGTTGCGGCTCGCGATCCTGATTTCGGTGGCCCTGCACG
AATGTGGTCACATGTGGGTCGCGCGCCGACCGGGATGAAGGTACGTCGCTATTTGTCGGCT
TTGGCCCCACGTTGTGGTCGACCCGGCGCGGCGAGACCGAATACGGTGTCAAAGCCGTTCCGC
10 TGGGCGGCTTCTGTGACATCGCCGGCATGACCCCGGTGCGAGGAACGACCCCGACGAACGTG
ACCGTGCGATGTACAAGCAGGCCACCTGGAAGCGGGTCGAGTGTTATTCGCCGGGCCCCGGAA
TGAACCTCGCTATCTGCCTGGTGCTGATCTATGCCATCGCGCTGGTCTGGGGGCTGCCTAACCT
GCATCCGCCAACCAAGGGCCGTAATCGGCGAAACTGGCTGCGTTGCACAGGAAGTGAGCCAGG
GCAAGCTCGAGCAGTGACCCGGGCCCGGTCCGGCGGCGCTGGCCGGAATTCGCTCCGGTGAC
15 GTCGTGGTCAAGGTGCGTGACACCCCGGTGTCCAGTTTCGACGAGATGGCCGCCGCGGTGCG
CAAGTCACACGGCAGCGTCCCGATCGTTGTCGAGCGTGACGGCACCGCGATTGTTACCTACGT
GGACATCGAATCCACCCAACGCTGGATCCCTAACGGGCAGGGCGGTGAGCTCCAGCCGGCAAC
GGTCGGTGCGATTGGGGTGGGCGCCGCCGGGTGCGGCCTGTGCGCTACGGCGTGTTCTCCG
CCATGCCGGCCACATTGCGGGTCACCGGCGACCTGACCGTGAGGTGGGCAAGGCGCTGGCC
20 GCCCTCCCGACCAAGGTAGGTGCGCTGGTGCGGGCGATCGGCGGCGGGCAGCGTGACCCGC
AGACGCCGATAAGTGTTGGTGGGCGCCAGCATCATCGGCGGCGACACCGTCGACCATGGGCTG
TGGGTGGCGTTCTGGTTCTTCTGGCCAGCTGAACCTCATCCTGGCTGCGATCAACCTGCTGC
CGTTGCTGCCGTTGATGGCGGCCATATTGCCGTGCGGGTGTTGAGAGGATCCGCAACATGG
TCCGGTCGGCTCGTGGAAGGTGGCGGCCGACCGGTGAATTACCTCAAACCTCTTGCCGGCGA
25 CCTATGTGGTCTTGGTTCTTGTGCTCGGGTACATGCTCTTGACCGTCACCGCCGACCTGGTCAA
CCCGATTAGGCTTTTCCAGTAG

>Rv2870c - TB.seq 3181770:3183077 MW:45324 >emb|AL123456|MTBH37RV:c3183077-3181767,
Rv2870c SEQ ID NO:112
30 GTGGCTACCGGTGGACGCGTCGTGATCCGGCGGCGCGGTGACAACGAGGTGGTGGCGCACAA
TGATGAGGTGACCAACTCGACCGACGGGCGCGCTGACGGCCGGTTGCGGGTGGTGGTGCTGG
GCAGTACCGGCTCGATCGGCACCCAGGCGCTTCAGGTGATCGCCGACAATCCGGACCGTTTCG
AGGTAGTCGGGCTGGCCGCTGGCGGCGCCCATCTGGACACGTTGCTGCGACAACGTGCGCAG
ACCGGGGTGACCAATATTGCCGTGCTGACGAGCACGCGGCGCAGCGGGTCGGCGACATCCC
35 CTACCACGGATCCGACGCCGCCACCGGGTGGTCGAGCAGACCGAGGCCGACGTCGTCCTCA
ATGCGCTGGTCGGCGCGTTGGGCCTGCGACCGACGTTGGCCGCGCTCAAGACGGGTGCCCGG
CTGGCGCTGGCCAACAAGGAATCGCTGGTCGCCGGTGGTTCGCTGGTGCTGCGGGCGGCGCG

GCCCGGTCAGATCGTGCCGGTCGACTCCGAACACTCCGCGCTGGCCAGTGCCTGCGCGGCG
GCACTCCCCGACGAGGTGCGCAAGCTGGTGCTGACGGCTCGGGAGGGCCGTTTCGGGGCTGG
TCCGCGGCCGACCTCGAGCATGTACCCCCGAGCAGGCTGGCGCGCATCCTACGTGGTCGATG
GGCCCGATGAACACGCTGAATTCCGGCGTCGCTGGTCAACAAGGGACTTGAGGTCATCGAAACC
5 CACCTGCTGTTCCGGCATCCCCTACGACCGCATCGATGTCGTGGTGCACCCCCAGTCGATCATCC
ATTCGATGGTCACCTTCATCGACGGTTCGACGATCGCCCAGGCCAGTCCCCCGGACATGAAGCT
ACCGATTTCTGTTAGCGCTGGGCTGGCCGCGTCGGGTGACGGCGCCGCTGCTGCCTGTGATTT
CCATACCGCGTCGAGCTGGGAGTTTCGAGCCGTTGGACACCGACGTCTTCCCCGCGGTGAGTT
GGCCCGGCAGGCCGGCGTAGCCGGTGGCTGCATGACCGCGGTTTACAATGCGGCGAACGAAG
10 AAGCAGCAGCGGCGTTCTTGCTGGCCGATCGGCTTCCCGGCCATCGTCGGCATCATCGCCG
ACGTGTTGCACGCTGCCGACCAATGGGCCGTCGAACCCGCTACCGTGGATGACGTACTCGACG
CGCAGCGCTGGGCCCGCGAGCGAGCGCAGCGCGCGGTATCTGGTATGGCTTCGGTGGCGATC
GCAAGCACGGCGAAGCCGGGCGCAGCGGGTCGACACGCATCGACGTTAGAAAGGTCCTGA

15 >Rv2922c smc member of Smc1/Cut3/Cut14 family TB.seq 3234189:3238055 MW:139610
>emb|AL123456|MTBH37RV:c3238055-3234186, smc SEQ ID NO:113
GTGGGTGCAGGGAGTCGGTTTCCGCTGGTGACCCGCTGCCGAGCGTTGGAGCTCGGCCTGA
CCGGTTACGCGGCCAACCACGCCGACGGACGCGTGTGGTGGTCGCCAGGGTCCGCGCGCT
GCGTGCCAGAAGCTGCTGCAGCTGCTGCAGGGCGACACGACACCGGGCCGCTCGCCAAAGT
20 CGTCGCCGACTGGTCGCAGTCGACGGAGCAGATCACCGGGTTCAGCGAGCGGTAATCTGGCC
CCTCGTGTAACCTCAAGAGTCTGACGTTGAAGGGCTTCAAGTCCTTCGCCGCGCCGACGACTTTA
CGCTTCGAGCCGGGCATTACGGCCGTCGTTGGGCCAACGGCTCCGGCAAATCCAATGTGGTC
GATGCCCTGGCGTGGGTGATGGGGGAGCAGGGGGCAAAGACGCTGCGCGGCGGCAAGATGG
AAGACGTCATCTTCGCCGGCACCTCGTCGCGTGCGCCGCTGGGCCGCGCCGAAGTCACCGTTA
25 GCATCGACAACCTCCGACAACGCACTGCCTATCGAATACACCGAGGTGTCGATCACCCGAAGAAT
GTTTCGCGACGGTGCCAGCGAATACGAAATCAACGGCAGCAGTTGCCGTTTGATGGATGTGCA
GGAGTTGCTGAGCGACTCCGGCATCGGCCGTGAGATGCATGTGATTGTTGGGCAAGGGAAGCT
CGAGGAGATCTTGCAGTCGCGGCCTGAGGATCGCGGGCGTTTCATCGAGGAAGCCGCGGTG
TGCTCAAGCATCGCAAGCGCAAGGAAAAAGCTCTGCGCAAACCTCGACACGATGGCGGCGAACC
30 TGGCCCCGGCTCACCGATCTGACCACCGAGCTCCGGCGTCAACTCAAACCGCTGGGCCGGCAG
GCCGAGGCGGCCAGCGTGCCGCGGCCATCCAAGCCGATCTGCGCGACGCCCGGCTGCGCCT
GGCGGCCGACGACTTGGTAAGCCGACAGGCCGAACGGGAAGCGGTCTTTCAGGCCGAGGCTG
CGATGCGCCGCGAGCATGACGAGGCCGCCCGCCGGCTGGCGGTGGCATCCGAGGAGCTGGC
CGCGCATGAGTCCGCGGTGCGCGAACTCTCGACGCGGGCCGAGTCGATCCAGCACACTTGTT
35 CGGGCTGTCTGCGCTGGCCGAACGGGTGGACGCTACGGTGCGCATCGCCAGCGAACGCGCCC
ATCATCTCGATATCGAGCCGGTAGCGGTGACGACACCGACCCAGAAAGCCGAGGAGCTAG
AAGCCGAGGCCAGCAGGTGGCCGTCGCCGAGCAACAACTGTTAGCGGAGCTGGACGCGGCG

CGTGCCCGACTCGATGCTGCCCCTGCAGAGCTGGCCGACCGGGAGCGCCGCGCCGCGGAGG
CCGACCGGGCACACCTGGCGGCGGTCCGGGAGGAGCGGACCGCCGTGAGGGACTGGCGCG
GCTGGCTGGCCAGGTGGAGACCATGCGGGCGCGTGTGCAATCGATCGATGAGAGCGTGGCAC
GGTTGTCCGAGCGGATCGAGGATGCCGCAATGCGCGCCAGCAGACCCGAGCCGAGTTTGAA
5 ACCGTGCAGGGCCGCATCGGTGAACTGGATCAAGGCGAGGTGCGCCTGGATGAGCACCACGA
GCGTACTGTGGCCGCGTTGCGGTTGGCCGACGAACGCGTCCGCGAGCTGCAATCCGCCGAAC
GCGCCGCCGAACGCCAGGTGGCATCGCTACGGGCTCGCATCGATGCGCTCGCAGTGGGGCTA
CAGCGCAAGGACGGCGCGGCGTGGCTGGCGCACAAATCGCAGTGGCGCAGGGCTTTTCGGTTC
GATCGCCCAATTGGTGAAGGTACGTTCCGGCTATGAAGCGGCACTGGCCGCGGCGCTCGGGC
10 CGGCGGGCCGACGCACTTGCGGTGGACGGCCTGACTGCCGCGGGTAGTGCCGTCAGCGCACTC
AAACAAGCCGACGGCGGTGCGCGGTCCTCGTGCTGAGTGAAGTGGCCGGCCCCGCAAGCCCC
CCAATCCGCCTCGGGGGAGATGCTGCCTAGCGGCGCCAGTGGGCCCTAGACCTGGTCGAGT
CTCCACCGCAGTTGGTTGGCGCGATGATCGCCATGCTTTCGGGTGTGCGGGTGGTCAACGACC
TGACTGAGGCAATGGGCCTGGTCGAGATTGTCGCGGAGCTACGCGCGGTACCGTTGACGGTG
15 ATCTGGTGGGCGCCGGCTGGGTGAGCGGCGGATCGGACCGCAAGCTGTCCACCTTGAGGTC
ACCTCCGAGATCGACAAGGCCAGGAGTGAGCTGGCCGCTGCCGAGGCGCTGGCGGCGCAATT
GAATGCGGCCCTGGCCGGTGGCTGACCGAGCAGTCCGCCCGCCAGGACGCGGCCGAGCAA
GCCTTGCGCGCGCTTAACGAATCCGACACGGCCATCTCGGCGATGTACGAGCAGCTGGGCCGC
CTCGGGCAGGAGGCCCGCGCGGCGGAAGAAGAGTGGAACCGGTTGCTGCAGCAGCGTACGGA
20 ACAGGAAGCCGTGCGCACACAGACTCTCGACGACGTCATACAACTTGAGACCCAGCTGCGTAA
GGCCCAGGAGACCCAACGGGTGCAGGTGGCCCAACCGATCGACCGCCAGGCGATCAGTGCCG
CTGCCGATCGCGCCCGCGGTGTGCAAGTGGAAGCCCGGCTGCGGGTGCGCACCGCCGAGGAA
CGCGCCAACGCGGTTGCGGGGCGGGCCGATTGCTGCGCCGTGCGGCTGCGGGCGGAACGTG
AGGCGCGGGTGCGGGCTCAGCAAGCACGCGCCGCAAGACTGCATGCGGCCGCGGTGGCCGC
25 AGCGGTGCGCGACTGCGGACGGCTGCTGGCCGGGCGGTTGCACCGGCGGTGGACGGGGCG
TCGCAACTGCGCGACGCGTCCGGCCGCGCAACGTGAGCAGCGTTAGCGGCGATGGCCGCGGT
GCGCGACGAGGTGAACACGCTGAGCGCCCGAGTGGGGGAACACCGATTGCTGCACCGCG
ACGAGCTGGCTAACGCGCAGGCGGCGCTGCGTATCGAGCAGCTTGAGCAGATGGTGCTAGAG
CAGTTCGGAATGGCGCCGCGGCGACTTGATCACCGAATACGGTCCACATGTGGCGCTACCACCG
30 ACCGAGCTCGAGATGGCTGAGTTCGAGCAAGCCCGCGAAGCGGCGAGCAGGTGATTGCGCC
CGCCCCCATGCCGTTGACCGGGTTACCCAGGAGCGCCGGGCCAAACGCGCCGAGCGTGCGC
TTGCCGAGTTGGGCAGGGTCAACCCGCTGGCGCTCGAAGAGTTTGCTGCCCTTGAGGAGCGCT
ACAATTTCTGTCCACCCAACTCGAGGATGTCAAGGCTGCCCGCAAGGATCTGCTGGGCGTCTG
CGCCGATGTTGACGCGCGCATCTGCAGGTGTTCAATGACGCGTTCGTAGACGTGGAACGCGA
35 ATTCGCGGCGTGTTACCGCATTGTTCCCCGGTGGTGAAGGACGGCTGCGGCTGACCGAGCC
CGACGACATGCTCACCACCGGCATCGAGGTGCAAGCCCGCCCGGGCAAGAAGATTACCC
GACTGTCTTTGCTCTCCGGTGGCGAGAAGGCGCTGACCGCGGTGGCGATGCTGGTCGCGATCT

TTCGTGCCCGTCCATCGCCGTTCTACATCATGGACGAGGTGGAGGCCGCCCTCGACGACGTGA
ACCTGCGCCGACTGCTCAGCCTGTTCTGAACAGCTGCGAGAGCAGTCGCAGATCATCATCAC
CCACCAGAAGCCGACGATGGAGGTGCGCGGACGCACTGTACGGCGTAACCATGCAGAACGACG
GCATCACCGCGGTCTCTCGCAGCGCATGCGCGGTGTCAGCAGGTGGATCAGCTGGTTACCAATT
5 CCTCGTAG

>Rv2925c mc RNase III TB.seq 3239829:3240548 MW:25400

>emb|AL123456|MTBH37RV:c3240548-3239826, mc SEQ ID NO:114

ATGATCCGGTCAAGACAACCCCTGCTCGACGCACTCGGTGTGGACCTCCCGGACGAGCTGCTC
10 TCACTGGCGTTGACCCACCGCAGCTACGCCTACGAGAACGGCGGGCTGCCGACCAACGAGCGT
TTGGAGTTTCTCGCGCATGCCGTGCTAGGGCTGACCATCACCGACGCGCTGTTCCATCGTCATC
CTGATCGGTGCGAGGGGGATCTGGCCAACTGCGGGCCAGCGTAGTCAACACCCAGGCCCTG
GCCGACGTGCGACGCCGCCCTCTGTGCGGAAGGCCCTCGGTGTTACGTGCTATTGGGTGCGGGC
GAGGCGAACACCGGCGGGGGCCGACAAGTCCAGCATTCTGGCCGACGGTATGGAATCGCTGCT
15 GGGCGCGATCTACCTGCAACACGGTATGGAGAAGGCCCGTGAGGTGATCCTGCGGCTGTTTGG
CCCGTTGCTGGACGCCGCGCCGACCCTGGGTGCGGGATTGGATTGGAAGACCAGCTTGCAGG
AGCTGACTGCAGCGCGAGGGCTGGGTGCGCCGTACATACCTGGTCACCTCCACCGGCCCGGAC
CACGATAAGGAATTCACCGCGGTGGTTGTCGTGATGGACAGCGAATACGGTTCAGGAGTGGGC
CGGTCCAAAAAGAACCCGAGCAAAAGCCGCGGCGGCCGCTTGGAAGCCCTGGAAGTGCTC
20 GACAACGCCATGCCGGGCAAAACCTCCGCCTAA

>Rv2934 ppsD TB.seq 3262245:3267725 MW:193317

>emb|AL123456|MTBH37RV:3262245-3267728, ppsD SEQ ID NO:115

ATGACAAGTCTGGCGGAGCGCGCGGCGCAACTGTGCGCGAACGCGCGAGCGGCCCTGGCGCG
25 CGAGCTCGTCCGTGCGGGTACGACCTTCCCGACCGACATCTGCGAGCCGGTGGCGGTGGTGG
GCATCGGCTGTCGCTTCCGGGGAATGTGACTGGGCCAGAGAGCTTTTGGCAGCTACTGGCCG
ACGGTGTGGACACAATCGAGCAGGTGCCGCCTGATCGGTGGGATGCGGACGCGTTCTACGATC
CCGATCCTTCGGCGTCGGGTGCGATGACGACGAAATGGGGTGGTTTCGTTTCCGATGTGACG
CGTTGACGCCGACTTTTTCGGAATCACTCCTCGGGAAGCCGTGGCGATGGACCCGCAGCATC
30 GGATGCTGCTCGAGGTTGCCTGGGAAGCGTTGGAGCACGCGGGTATTCCGCCGGATTCCCTGA
GCGGCACTCGAACCGGCGTGATGATGGGTCTGTGTCGTGGGACTACACGATCGTCAATATCG
AGCGCAGAGCCGACATCGACGCGTACCTGAGCACCGGAACCCCGCACTGTGCCGCGGTGGGG
CGGATCGCGTATCTGTTGGGATTGCGTGGTCCGGCGTCGCCGTAGATACCGCTTGTTGTCGT
CGCTGGTGGCAATTCATTGGCGTGTGAGAGCCTTCGCCTGCGTGAAACCGACGTGGCATTGG
35 CGGGCGGGGTGCAGCTCACCTTGTACCGTTACCGCCATCGCGCTGTCCAAGTGGTCCGGCG
TGTCACCGACCGGCCGATGCAACAGCTTCGACGCCAACGCGGATGGATTGTCGCGGGCGAG
GGCTGCGGCGTGGTGGTGCTCAAGCGGTTGGCCGACGCGGTGCGCGACCAGGACCGGGTGCT

TGCGGTGGTCCGCGGTTCCGGCAACTAACTCCGATGGTCCGATGACCGCACCGAA
CGCGCTGGCGCAGCGTGACGTGATCACATCCGCCCTCAAGCTTGCGGATGTTACCCCTGACAG
CGTGAAGTATGTCGAAACACACGGCACCGGAACGGTGTGGGGACCCCATCGAGTTCGAGTC
GCTGGCGGCCACTTATGGCCTGGGTAAAGGCCAGGGCGAGAGCCCGTGCGCATTGGGGTCGG
5 TCAAGACCAACATCGGCCACCTGGAGGCGGGCCGCGGTGTGGCTGGATTCATCAAGCGGTGC
TGGCGGTGCAACGTGGGCACATTCCCCGCAACTTGCACTTCACCCGGTGGAACCGGGCCATCG
ACGCGTCGGCGACGCGGCTGTTCTGTGCCGACCGAAAGCGCCCCGTGGCCGGCGGCTGCCGGT
CCACGCAGGGCTGCGGTGTCATCGTTCGGCCTCAGCGGGACCAACGCGCACGTGGTGGTGA
GCAGGCACCCGACACCGCAGTAGCCGCAGCCGGCGGCATGCCGTATGTTTCGGCGCTGAACG
10 TCTCCGGCAAGACGGCCGCGCGGGTGGCGTCCGGCGCGCGGTGCTGGCCGACTGGATGTC
GGGGCCGGGCGCGCGGCACCACTGGCCGACGTGGCACACACGTTGAACCGGCACCGGGCC
CGGCACGCCAAGTTCGCCACCGTCATCGCGCGTGACCGCGCCGAGGCGATCGCGGGGTTGCG
AGCGCTGGCGGGCCGGACAACACGCGTTGGGGTGGTGGATTGCGACCAGCATGCCGGTGGGC
CTGGCCGGGTTTTTGTGTATTGGGTGAGGGCTCGCAGTGGGCGTCGATGGGCCAGCAGTTGC
15 TGGCCAACGAACCGGCGTTCGCCAAGGCGGTAGCCGAGCTGGATCCGATATTCGTTGACCAGG
TTGGCTTTTCGCTGCAGCAAACGCTTATCGACGGCGACGAGGTGGTGGGCATCGACCGCATCC
AGCCGGTGCTGGTGGGATGCAGTTGGCGCTGACCGAGTTATGGCGGTCCTATGGGGTGATT
CAGATGCCGTGATCGGGCACTCGATGGGTGAGGTGTCGGCGGCAGTGGTGGCCGGCGCGTTG
ACGCCCCGAGCAGGGCTTGCGGGTCATCACCAACCGGTGCGCGTTGATGGCGCGGCTGTCCGG
20 GCAGGGAGCGATGGCGCTGCTCGAGCTGGATGCCGACGCCGCCGAGGCGCTGATTGCCGGCT
ATCCGCAGGTGACGCTGGCGGTGCATGCGTCACCGCGCCAGACGGTGATCGCCGGGCGCGCC
GAGCAGGTGGACACGGTGATCGCGGCGGTAGCGACGCAAAACCGGTTGGCGCGCCGCGTCGA
AGTCGACGTGGCCTCCCATCACCCGATCATCGATCCATACTGCCCGAGTTGCGAAGCGCGTTA
GCGGATTTGACTCCGCAGCCGCCGAGCATCCCGATCATTTCCACTACGTACGAAAGCGCGCAG
25 CCGGTGGCGGATGCCGACTATTGGTCCGCCAACCTGCGCAACCCGGTGCGATTCCACCAGGCC
GTCACCGCCGCGGGTGTGACACACAACACCTTCATCGAAATCAGCCCTCACCCCGTGCTCAG
CACGCACTCACCGACACCCCTGGATCCGGACGGCAGCCATACAGTCATGTCGACGATGAACCGC
GAACTGGACCAGACGCTGTATTTCCACGCCCAACTCGCCGCGGTGCGTGTGGCTGCGTCCGAG
CACACCACCGGTGCGCTTGTGACCTGCCCCCACACCGTGGCACCATCAGCGATTCTGGGTC
30 ACGGATCGTTCCGCGATGTCCGAGCTGGCCGCGACCCACCGCTCCTGGGCGCGCACATCGA
GATGCCGCGCAACGGAGACCATGTCTGGCAGACCGATGTCGGCACCGAGGTCTGTCCCTGGTT
GGCAGACCACAAGGTGTTTCGTCACCCATCATGCCGGCCGCGGGGTTCCGCCGAGATCGCCTT
GGCGGCGGGCAGCGAAGCCCTCGGCACAGCCGCCGACGCCGTGCGACCCAACATCGTGATCA
ACCAGTTCGAGGTGGAGCAGATGCTGCCCCCTGACCGGCCACACGCGCTAACGACGCAGTTAA
35 TTCGCGGCGGGGACAGCCAGATTGGGTCGAGATCTATTCCCGCACGCGTGGCGGAGAGTTCT
GCCGACACGCCACGGCCAAGGTTGAACAATCGCCGCGGAATGTGCGCACGCGCACCCGGAA
GCCCAAGGTCCCGCCACCGGGACAACAGTGTGCCGGCGGATTTTTATGCCCTGCTCCGCCAA

ACCGGCCAACACCATGGTCCGGCGTTCCGCGCCTTAAGCCGGATCGTGCGCCTGGCCGATGGT
TCCGCGGAAACCGAGATCAGCATTCCCGACGAGGCGCCGCGCCATCCCGGGTATCGGCTGCA
CCCCGTGGTATTGGATGCGGCATTGCAAAGCGTGGGTGCCGCGATACCCGACGCGGAGATCGC
GGGGTCGGCGGAAGCCAGCTATCTGCCAGTGTCTGTCGAGACCATCCGGGTGTACCGCGACAT
5 CGGTCCGACACGTACAGGTGTCTGTCGCCACCTGACAAACCTCGACGGCGGCACCGGAAAGATGG
GCAGGATCGTCCTAATCAACGACGCCGGCCACATAGCGGCCGAAGTGGACGGCATCTATCTGC
GTCGTGTCTGAACGCCGTGCGGTACCCCTGCCACTAGAGCAGAAGATCTTCGATGCCGAATGGA
CCGAAAGCCCGATCGCAGCCGTGCCGGCTCCGGAGCCAGCTGCCGAGACGACGCGGGGAAGT
TGGCTGGTACTCGCCGATGCAACGGTGGATGCGCCAGGCAAGGCCCAAGTCGATGGC
10 CGACGACTTCGTGCAGCAGTGGCGCTCACCGATGCGGGCGGGTGACACCGCCGATATCCACGA
CGAATCGGCGGTGCTGGCCGATTTGCAGAAACGGCAGGCGATCCCGAGCACCCGCCGGTTG
GCGTGGTGGTGTCTGTCGGCGGTGCCTCGAGTCGACTGGACGACGAGCTGGCGGCGGGCGCGC
GACACGGTGTGGTTCGATCACCACGGTGGTTCTGTCGGTCTCGGCACGTGGCACGGCCGATCA
CCGCGGCTATGGTGGTACCCGGGGGCGGACTTTCCGTTGCCGACGACGAGCCGGGAACACC
15 CGCGGCGGCTTCCTTGAAAGGGCTGGTGCGGGTGTCTGCCTTCGAGCACCCGGACATGCGCA
CCACCCTGGTTCGATCTGGACATCACACAAGACCCGCTGACCGCGCTGAGCGCGGAACCTGCGGA
ATGCCGGGAGTGGGTGCGCCATGATGACGTGATCGCGTGGCGCGGCGAGCGCAGGTTCTGTC
GAACGGCTGTCTGCGCGCCACGATCGATGTATCCAAAGGGCATCCGGTGGTGCGCCAGGGAGC
GTCGTACGTCTGACCCGGCGGCTCGGCGGTCTCGGCCTGGTCTGTCGCTCGTTGGCTGGTGG
20 ACCGCGGCGCGCGGCGGGTGGTGTCTGGGTGGCCGCGAGCGATCCCACTGACGAGCAGTGCAAC
GTCCTGGCCGAACTGCAGACCCGCGCCGAGATCGTGGTTGTCCGTGGCGACGTGGCATCGCC
GGGGGTGGCAGAAAAGCTGATTGAGACGGCCCGACAGTCTGGGGGCCAATTGCGCGGCGTCTG
TGCACGCCGCCGCGGTCTCGAAGACAGCCTGGTGTCTCTATGAGCAGGGACAACCTAGAAC
GGGTGTGGGCACCCAAGGCCACCGGTGCGCTGCGCATGCACGAAGCCACCGCTGACTGCGAG
25 CTCGACTGGTGGCTCGGATTCTCTCCGCGCTTCGCTATTGGGTTCTCCCGGGCAAGCGGCCT
ACGCGTGCGCCAGCGCGTGGCTGGACGCGCTGGTCTGGATGGCGCAGGGCATCCGGCCTGCC
GGCCGCGGTGATCAACTGGGGTCCGTGGTTCGGAGGTAGGCGTCGCCAGGCCTTGGTGGGCA
GTGTTCTCGACACGATCAGTGTCTGAGAAGGCATCGAGGCTCTCGACTCATTGCTTGCCGCCGA
CCGGATCCGCACTGGAGTGGCTCGGCTGCGTGCCGATCGGGCCCTGGTTCGATTCCCGGAGA
30 TCCGCAGCATCAGCTACTTCACCCAGGTGGTTCGAGGAGCTGGACTCGGCGGGTGACCTCGGCG
ACTGGGGCGGGCCCGACGCGCTTGCCGACCTCGACCCGGGCGAGGCGCGGCGCGCGGTGAC
CGAGCGGATGTGTGCGCGCATCGCTGCGGTGATGGGCTACACTGACCAGTCGACTGTGCAACC
CGCCGTGCCCTTGGACAAGCCCCTGACCGAGCTGGGGCTGGATTCTCTGATGGCGGTACGAAT
ACGCAACGGCGCGCGGGCGGATTTGGCGGTGGAACCGCCGGTAGCGCTGATACTGCAAGGCG
35 CGTCTTGTCATGACCTGACGGCGGACTTAATGCGCCAACTCGGGCTCAATGATCCCGATCCGG
CGCTCAACAACGCTGACACTATTGCGGACCGGGCGCGCCAGCGCGCGGCAGCGCGACACGGA
GCCGCGATGCGGCGCCGACCTAACCTGAAGTACAGGGAGGATAA

>Rv2946c pks1 TB.seq 3291503:3296350 MW:166642

>emb|AL123456|MTBH37RV:c3296350-3291500, pks1 SEQ ID NO:116

GTGATTTTCGGCGAGATCGGCTGAGGCGTTGACGGCGCAGGCGGGTCGACTTATGGCCACGTG
5 CAGGCCAACCCAGGGCTGGATCCGATCGATGTGGGGTGCTCGTTGGCCAGTCGCTCGGTGTTT
GAGCACCGAGCGGTGGTGGTCGGCGCAAGCCGTGAGCAACTGATTGCCGGGCTGGCTGGGCT
CGCGGCGGGCGAGCCGGGTGCCGGCGTGGCGGTGCGTCAGCCAGGGTCGGTGGGCAAGACG
GTGGTCGTGTTTCTGGGCAGGGCGCGCAGCGCATCGGGATGGGCCGCGAGTTGTACGGCGA
GTTGCCCGTGTTTGGCGAGGCATTGATGCGGTGGCCGACGAGTTGGACCGGCATCTGCGGTT
10 GCCGCTGCGCGACGTATTTGGGGTGCCGATGCGGATTTGCTTGACAGCACCGAATTTGCTCAG
CCCGCGTTGTTGCGGGTGGAGGTGGCATCGTTCGCGGTGTTGCGGGATTGGGGTGTGCTTCCG
GACTTCGTGATGGGTCACTCCGTTGGAGAGCTGGCGGCGGCGCACGCGGCCGGTGTGTTGAC
GTTGGCGGACGCGGCGATGCTGGTGGTGGCGCGGGGCCGGTTGATGCAGGCGCTGCCGGCA
GGCGGTGCGATGGTGGCGGTGGCTGCCAGTGAGGACGAGGTGGAGCCGCTGCTGGGTGAGG
15 GTGTGGGGATCGCTGCGATCAACGCGCCCGAATCGGTGGTGTGATCTCCGGTGCGCAGGCCGCG
GCAAATGCGATTGCGGATCGGTTGCGCGCGCAGGGTCGCGCGGTGCACCAAGTTGGCGGTCTC
GCATGCGTTTCATTGCGCGTTGATGGAGCCGATGCTCGAGGAGTTGCGCGGTGTCGCGGCCCG
GGTGCAGGCACGCGAGCCCCAGCTTGGGCTGGTGTGCAACGTGACGGGCGAGTTGGCCGGCC
CTGATTTGCGGTGCGCGCACTACTGGGTGGACCACGTTGCTCGGCCGGTGCGCTTCGCGGACA
20 GTGCGCGTCATTTGCAGACCCCTTGGGGCGACCCACTTCATCGAGGCCGGCCCGGGAAGTGTT
TGA CTGGCTCGATCGAGCAGTCCTTGCCCCGGCTGAGGCGATGGTGGTGTGATGCTGGGCA
AAGACCGGCCCGAGCTGGCCTCGGCGCTCGGTGCTGCCGGTCAGGTGTTACCAACCGGTGTG
CCGGTGCACTGGTGGCGGGTGTTCGCGGGCTCGGGTGACGGCGGGTGCACTGCCACGTA
TGCGTTTCAGCGACGGCGGTTTTGGGAGACGCCGGGCGCGGATGGGCCCGCGATGCGGCCG
25 GGTGGGTCTGGGCGCGACCGAGCATGCCTTGTGGGTGCGGTGGTGCAGCGGCCCGATTCT
GACGAGGTGGTGTGACCGGCCGGTGTGCTGCTGCGGATCAGCCGTGGCTGGCCGACCACGT
GGTGAACGGGGTGGTGTGTTCCCCGGGGCGGGTTTTGTGGAGTTGGTGTGATCCGCGCCGGTG
ATGAGGTGCGGTGCGCGCTCATCGAAGAGTTGGTGTGCGCCGACCGTTGGTGTGACCCGG
GTGTGCGGGGTTCAAGTGCAAGGTGGTGTGCGGGGCTGCCGATGAATCCGGGCACCGTGCGGTG
30 TCGGTGTATTCGCGGGTGATCAATCCAGGGTTGGTTGCTGAACGCCGAAGGCATGCTGGGG
GTGGCTGCCGCTGAGACGCCGATGGATTTGTCCGTGTGGCCGCCCGAGGGCGCGGAGAGTGT
GGATATCTCGGACGGCTATGCGCAGTTGGCCGAGCGCGGTTATGCCTACGGCCCCGCGTTTCA
GGGTCTGGTGGCGATCTGGCGGCGGGGGTGGAGCTGTTGCGCGAAGTTGTAGCCCCCGGCG
AGGCCGGCGTGGCCGTGACCGAATGGGGATGCATCCGGCGGTGTTGGACGCGGTGCTGCAT
35 GCCCTCGGGCTGGCCGTGAGAAAGACCCAGGCGAGCACCGAGACGAGACTGCCGTTTTGCTG
GCGTGGGGTGTGCTGCATGCCGCGGGCGCTGGACGGGTGCGGGCCCGCTTCGCGTCCGCG
GGCGCGGATGCGATTTCCGTGGACGTCTGCGACGCCACTGGGCTGCCGGTGTGACGGTGCG

CTCGCTGGTTACTCGCCCGATAACCGCAGAACAGCTGCGCGCCGCGCTGACCGCGGCCGGCG
GTGCGTCCGATCAGGGGCGCTGGAAGTGGTGTGGTCGCCGATCTCGGTGGTCAGCGGCGGC
GCTAACGGGTCCGCCCCACCTGCCCCGGTGTCTTGGGCGGACTTTTGCGCCGGCAGTGATGGT
GACGCCAGTGTCTGTGGTGTGGGAAGTCTGAGTCTGCCGGTGGCCAAGCATCCTCGGTGGTGGG
5 CTCGGTGTATGCBGCCACCCACACCGCCCTGGAGGTGTTGCAGTCCTGGCTCGGCGCBGATCG
GGCGGCCACGTTGGTGGTGTGACCCATGGTGGCGTGGGGCTGGCTGGCGAGGACATCAGCG
ACCTGGCCGCCGCCGCGGTGTGGGGCATGGCGCGTTCCGCGCAGGCCGAAAATCCCGGCCG
GATCGTGTGATCGACACCGATGCGGCGGTGGATGCCTCGGTGCTAGCCGGCGTCGGGGAAC
CCCAGCTGCTGGTGC GCGCGGCACTGTGCACGCCCCCGGCTGTCCCCGGCCCCGGCGTTG
10 CTAGCGTTACCGGCGGCAGAGTCGGCGTGGCGATTGGCCGCCGGTGGTGGCGGGACCCCTGGA
GGATTTGGTGATCCAGCCCTGCCCGGAGGTACAGGCACCGCTACAGGCGGGGACAGGTGCGCG
TGGCGGTGGCGGCCGTCGGGGTCAACTTCCGCGATGTGGTGGCCGCCCTAGGGATGTATCCC
GGCCAGGCCCCACCGCTGGGTGCCGAAGGCGCCGGGTGGTGGCTTGAGACCGGTCCCGAAGT
GACCGATCTTGCCGTCGGTGACGCCGTGATGGGATTCTGGGCGGGGCGCGTCCGCTGGCGG
15 TGGTGGATCAGCAACTGGTTACCCGGGTGCCGCAAGGCTGGTCGTTTGCTCAGGCAGCCGCTG
TGCCGGTGGTGTCTTGACGGCTGGTACGGGTGGCCGATTTAGCCGAGATCAAGGCGGGCG
AATCGGTGCTGATCCATGCCGGTACCGCGGTGTGGGCATGGCGGCTGTGCAGCTGGCTCGC
CAGTGGGGCGTGGAGGTTTTCTGCACCGCCAGCCGTGGCAAGTGGGACACGCTGCGCGCCAT
GGGGTTTGACGACGACCATATCGGCGATTCCCGCACATGCGAGTTCGAGGAGAAGTTCCTGGC
20 GGTCACCGAGGGCCGCGGGGTTGATGTGGTGTCTGACTCGCTGGCCGGTGAGTTCGTGGATG
CGTCGCTGCGCTTACTGGTCCGCGGTGGGCGTTTCTCGAGATGGGCAAGACGGATATCCGCG
ATGCGCAGGAGATCGCCGCTAATTATCCCGCGGTGCAGTATCGGGCGTTCGACCTGTGCGAGG
CCGGCCCCGGCACGCATGCAGGAGATGTTGGCCGAGGTGCGGGAGCTGTTGACACCCGGGAG
CTGCACCGGCTACCGGTACCCACGTGGGATGTGCGCTGCGCCCCGGCGGCCCTCCGGTTCATG
25 AGCCAGGCCCGCCATATCGGCAAGGTTGTCTTAACCATGCCCTCGGCGTTGGCCGACCGGCTT
GCCGACGGCACGGTGGTGATCACCGGTGCCACCGGGGCGGTTGGTGGGGTGTGGCCCCGCCA
CCTGGTTGGCGCCTATGGGGTGCCTCATCTGGTGTGGCCAGTCGGCGGGGCGATCGCGCGG
AGGGAGCGGCCGAATTGGCCGCCGACTTGACGGAGGCCGGCGCCAAGGTGCAGGTGGTGGC
CTGTGACGTGGCCGATCGCGCTGCGGTAGCGGGGTTGTTGCCAGCTGTGCGGGGAGTACCC
30 GCCGGTGC GCGGGGTGATTCATGCCGCCGGCGTGCTCGATGACGCAGTGATCACCTCGTTGAC
ACCGGACCGCATCGATACGGTGTGCGGGGCCAAGGTGGACGCGGCGTGGAACCTGCACCAGG
CCACCACTGACCTGGATTTGTCGATGTTTGCGCTGTGCTCATCGATCGCGGCCACGGTCCGGCTC
GCCGGGGCAGGGCAACTACTCGGCGGCAAACGCGTTTCTGGACGGGTGGCCGCTACCCGGC
AGGCCGCAGGGTTGGCCGGGATATCACTGGCGTGGGGTTGTGGGAACAGCCTGGCGGCATG
35 ACCGCGCATTTGAGCAGCCGAGATCTGGCCCGCATGAGCCGCAGCGGGCTGGCTCCGATGAG
CCCTGCCGAAGCGGTGGAATTGTTGACGCTGCGCTGGCCATCGATCACCTCTGGCGGTGGC
CACGCTCTTGACCGGGCTGCACTAGACGCCCGGGCCAGGCCGGTGC GTTGCCGGCGCTGT

TCAGCGGGCTCGCGCGCCGCCACGCCGACGCCAAATCGACGACACCGGTGACGCCACCTCG
TCGAAGTCGGCGCTGGCTCAACGCCTACACGGGCTGGCCGCGGACGAACAACCTCGAGCTGCTA
GTGGGGCTGGTGTGTCTGCAGGCAGCGGCAGTGCTGGGTAGGCCCTCCGCCGAGGACGTGCA
CCCCGACACCGAATTCGGCGACCTCGGTTTTCGACTCATTAACGGCTGTGGAGTTACGCAACCGC
5 CTCAAACCGCCACCGGACTGACGCTGCCACCTACCGTGATTTTCGATCATCCCACTCCCACTG
CGGTGCGCGAGTATGTCGCCCAGCAAATGTCTGGCAGCCGCCAACGGAATCCGGTGATCCGA
CGTCGCAGGTTGTGCAACCCGCCGCCGCGGAAGTATCGGTCCATGCCTAG

>Rv3014c ligA DNA ligase TB.seq 3372545:3374617 MW:75258

10 >emb|AL123456|MTBH37RV:c3374617-3372542, ligA SEQ ID NO:117
GTGAGCTCCCCAGACGCCGATCAGACCGCTCCCCAGGTGTTGCGGCAGTGGCAGGCACTGGC
CGAGGAGGTGCGTGAGCACCAGTTCCGTTATTACGTGCGGGACGCGCCGATCATCAGCGACGC
GGAATTCGACGAGCTGCTGCGCCGTCTGGAAGCCCTCGAGGAGCAGCATCCCCGAGCTGCGCA
CGCCCCGATTCGCCGACCCAGCTGGTCGGCGGTGCCGGCTTCGCCACGGATTTGAGCCCCGTC
15 GACCATCTCGAACGAATGCTCAGCCTCGACAACGCGTTCACCGCCGACGAACTCGCCGCCTGG
GCCGGCCGCATCCATGCCGAGGTGCGGAGACGCCGCACATTACCTGTGTGAGCTCAAGATCGAC
GGCGTCGCGCTGTCTTTGGTCTACCGCGAGGGACGGCTGACCCGGGCCTCCACCCGCGGCGA
CGGGCGCACCGGCGAGGACGTACCCTGAACGCCCGGACCATCGCCGACGTTCCCGAACGGC
TCACCCCCGGCGACGACTACCCGGTGCCCGAGGTCTCGAGGTCCGCGGGCAGGTCTTCTTCC
20 GGCTGGACGACTTCCAGGCGCTCAACGCCAGCCTCGTCGAGGAGGGCAAGGCGCCGTTGCC
AACCCCCGCAACAGCGCGGGGATCGCTGCGCCAGAAAGACCCGGCGGTACCGCGCGCCG
CCGGCTGCGGATGATCTGCCACGGGCTGGGCCACGTGGAGGGCTTTCGCCCGGCCACCTGC
ATCAGGCATACCTGGCGTTGCGGGCATGGGGACTGCCGGTTTCCGAACACACCACCTGGCAA
CCGACCTGGCCGGTGTGCGCGAGCGCATCGACTACTGGGGCGAGCACCGCCACGAGGTGGAC
25 CACGAAATCGACGGCGTGGTGGTCAAAGTCGACGAGGTGGCGTTGCAGCGCAGGCTGGGTTC
CACGTGCGGGGCGCCGCGCTGGGCCATCGCTACAAGTACCCGCCCGAGGAAGCGCAGACCA
AGCTGCTCGACATCCGGGTGAACGTGCGCCGACCGGGCGGATCACGCCGTTTTCGTTTCATGA
CGCCGGTGAAGGTGGCCGGTTCGACGGTGGGACAGGCCACCTGCACAACGCCTCGGAGATC
AAGCGCAAGGGCGTGCTGATCGGCGACACCGTGGTGATCCGCAAGGCCGCGACGTGATCCC
30 CGAGGTGCTGGGACCCGTCGTCGAACTGCGCGATGGCTCCGAACGCGAATTCATCATGCCAC
CACCTGCCCCGAGTGCGGTTCCCGTTGGCGCCGAGAAAGGAAGGCGACGCCGACATCCGTT
GCCCCAACGCCCGCGGCTGCCCGGGCAACTGCGGGAGCGGGTTTTCCACGTGCCAGCCGC
AACGGCCTAGACATCGAGGTGCTCGGTTACGAGGCGGGTGTGGCGCTCTTGACGGCGAAGGT
GATCGCCGACGAGGGCGAGCTGTTGCGGCTGACCGAGCGGGACTTGCTGCGCACCGACCTGT
35 TCCGAACCAAGGCAGGCGAACTGTGCGCCAACGGCAAACGGCTGCTGGTCAACCTCGACAAGG
CCAAGGCGGCACCGCTGTGGCGGGTGTGGTGGCGCTGTCCATCCGCCATGTGGGGCCGACG
GCGGCCCCGCGCCCTGGCCACCGAGTTCGGCAGCCTTGACGCCATCGCCGCGGCGTCCACCGA

CCAGCTGGCCGCCGTGAGGGGGTGGGGCCGACCATTGCCGCCGCGGTACCGAGTGGTTCG
CCGTGCGACTGGCACC GCGAGATCGTCGACAAGTGGCGGGCCGCGGGGTGCGAATGGTGCAC
GAGCGTGACGAGAGTGTGCCACGCACGCTGGCCGGGCTGACCATCGTGGTCACCGGCTCGCT
GACCGGTTTTCTCCCGCGACGACGCCAAGGAGGCGATCGTGCCCCGCGGCGGCAAGGCCGCCG
5 GCTCGGTGTGCAAGAAGACCAACTATGTCGTCGCCGGAGACTCGCCGGGATCCAAATACGACA
AGGCGGTGGAGTTGGGGGTGCCGATTCTGGACGAGGATGGGTTCCGGAGACTGCTGGCCGAC
GGACCCGCGTCACGAACGTAA

>Rv3025c - NifS-like protein TB.seq 3383885:3385063 MW:40948

10 >emb|AL123456|MTBH37RV:c3385063-3383882, Rv3025c SEQ ID NO:118
ATGGCCTACCTGGATCACGCTGCCACCACCCCGATGCACCCCGCCGCCATCGAGGCGATGGCG
GCCGTGCAGCGCACCATCGGCAATGCGTCGTCGCTGCACACCAGCGGGCGCTCGGCCGCCCG
GCGGATCGAGGAGGCCCGTGAGCTGATCGCGGACAAGCTAGGCGCTCGTCCGTCCGAGGTGA
TCTTACCCGCGGGCGGGCACCGAAAGCGACAACCTGGCTGTCAAAGGTATCTATTGGGCACGCC
15 GCGATGCGGAGCCGCACCGCCGTGCGATCGTCACCACCGAGGTGGAACACCACGCCGTA CTG
GACTCGGTGAACTGGCTCGTGGAACACGAAGGCGCCCATGTGACCTGGGTGCCGACCGCCGC
CGACGGCTCGGTGTGCGCAACTGCGCTGCGCGAGGCACTGCAGAGCCACGACGACGTGCGCG
TGGTATCGGTGATGTGGGCCAACAACGAGGTGCGAACTATTCTACCGATCGCCGAAATGTCA GT
TGTCGCCATGGAATTCGGCGTGCCGATGCACAGTGATGCCATT CAGGCGGTGGGACAGCTCCC
20 GCTTGACTTCGGGGCCAGCGGGCTGTGCGCGATGAGCGTGCGCGGGCACAAATTCGGTGGCC
CGCCAGGAGTGGGTGCGTTGCTGCTGCGCCGCGACGTACCTGCGTGCCCCTTATGCACGGC
GGTGGGCAGGAGCGCGATATTCGTTCCGGCACACCCGATGTGCGCCAGTGCA GTTGAATGGCG
ACGGCCGCGCAGATCGCGGTGGACGGACTCGAGGAAAACAGCGCGCGGTTACGGCTGCTGCG
GGATCGTCTGGTCGAGGGTGTGCTGGCTGAGATTGACGATGTTTGCTTAACGGCGCCGATGA
25 CCCGATGCGGCTAGCGGGTAACGCGCACTTCACTTTCGTTGGCTGCGAAGGCGATGCGCTGTT
GATGTTGTTGGACGCTAACGGAATCGAGTGCTCAACCGGATCGGCCTGCACGGCAGGTGTAGC
GCAGCCCTCGCATGTGTTGATTGCAATGGGCGTCGACGCGGCCAGCGCCCGCGGATCATTGCG
TCTCTCGCTGGGGCACACCAGTGTTGAGGCTGATGTCGATGCCGCGTTGGAGGTGCTTCCCGG
GGCGGTGGCACGTGCACGGCGGGCCGCCCTAGCCGCGCGGGAGCATCCCGATGA

30 >Rv3080c pknK serine-threonine protein kinase TB.seq 3442656:3445985 MW:119420
>emb|AL123456|MTBH37RV:c3445985-3442653, pknK SEQ ID NO:119

ATGACCGACGTTGATCCGCACGCGACGCGGGGACCTGGTCCCGAATATCCCGCGGAACTG
CTTGAGGCTGGATTGACAATGTGAGGAGATCGGGCGCGGCGGATTCGGCGTCGTCTACCGC
35 TGCGTCCAGCCCTCGCTGGACCGCGCCGTCGCCGTCAAGGTATTGAGCACCGACCTGGATCGG
GACAATCTCGAGCGCTTCCTGCGCGAGCAGCGGGCCATGGGCCGCCCTTCCGGGCACCCGCA
CATCGTGACCGTCTTGCAAGGTGGGCGTGTGGCGGGTGGGCGGCCCTTCATCGTGATGCCCTA

CCACGCCAAGAATTCGTTGGAGACGCTGATTGCCGGGCACGGGCCGCTGGACTGGCGCGAGA
CGCTGTCGATCGGCGTCAAGCTCGCGGGAGCGCTGGAAGCCGCGCATCGCGTCGGCACCCCTG
CACCGTGACGTGAAGCCGGGGAATATCCTGCTGACCGACTACGGGGAACCGCAGCTGACCGAT
TTCGGAATCGCCAGAATCGCCGGGGGTTTCGAGACGGCGACCGGGGTGATTGCCGGTTCCCCG
5 GCTTTCACCGCGCCGGAAGTTCTCGAAGGAGCATCGCCGACGCCCGCCTCTGACGTGTACTCC
CTGGGCGCGACGTTGTTCTGTGCGCTGACCGGCCATGCCGCCTACGAGCGCCGCAGCGGTGA
GCGGGTGATCGCCAGTTCTGCGGATCACCTCGCAGCCGATCCCCGACCTGCGGAAGCAGG
GACTGCCCGCGGACGTGGCCGCGCCATCGAACGGGCGATGGCCCGCCATCCGGCGGATCGT
CCCCGACCGCGGCAGACGTTGGCGAGGAGCTTCGCGACGTTACGCGCCGCAACGGCGTCAG
10 CGTCGACGAGATGCCCTCCCCGTCGAGCTGGGCGTGGAACGCCGACGCTCGCCCCGAGGCGC
ACGCGGCGCATCGGCATACCGGCGGCGGCACCCCGACGGTCCCGACGCCTCCGACACCCGCG
ACCAAGTACCGGCCGTCGGTGCCACCGGCTCGCTGGTCACCCGCAGCCGGCTCACCGACAT
CCTGCGCGCCGGCGGACGGCGCCGGCTGATCCTCATCCACGCGCCCTCGGGATTCCGCCAAAA
GCACCCTGGCGGCGCAATGGCGGGAAGAGCTCTCGCGCGACGGCGCCGCGGTGCGCTGGCT
15 GACAATCGACAACGACGACAACAACGAGGTGTGGTTCTTGTGCGACCTGCTCGAGTCGATCCG
GCGGGTCCGGCCACGCTGGCCGAGTCGTTGGGGCACGTGCTCGAAGAGCATGGGGATGACG
CCGGCCGCTACGTGTTGACTTCGCTGATCGACGAAATCCACGAAAACGACGACCGGATCGCGG
TGGTGATCGACGACTGGCATCGGGTGTCGACAGCCGCACCCAAGCTGCCCTGGGTTTCTGCG
TGGACAACGGATGTCACCACTGCAGCTCATCGTGACCAGCTGGTCTCGCGCCGGTTTGCCGG
20 TGGGCAGGTTGCGGATCGGCGACGAAGTAGCCGAGATCGATTGCGCTGCTTTGCGCTTCGATA
CCGACGAGGCGCGCGCTGCTGAACGATGCTGGTGGTCTGCGATTGCCGCGCGCAGACGTG
CAGGCGCTGACTACCTCTACCGACGGGTGGGCCGCGGCGCTGCGGCTGGCCGCGCTGTCGCT
GCGCGGCGGGGGCGACGCGACCCAACCTGCGCGGACTTTCGCGCGCCAGTGACGTGATCC
ACGAATTCCTGAGCGAAAACGTGCTGGACACCCTGGAACCCGAACTGCGCGAATTCCTACTGGT
25 GGCATCGGTCACCGAACGCACGTGCGGCGGGCTGGCCTCGGCGCTGGCCGGGATCACCAATG
GGCGGGCGATGCTGGAAGAGGCCGAGCACCGCGGCTTGTTCTGCAACGGACCGAAGACGAC
CCGAATTGGTTTCGCTTCCACCAAATGTTCCGCGACTTTCCTCCACCGTCGCTCGAACGTGGCG
GGTCGCACCGGGTGGCGGAAGTGCACCGCAGGGCATCGGCCTGGTTCGCGGAGAACGGCTAC
CTGCACGAAGCCGTCGACCATGCACTGGCCGCGGGCGATCCCGCGCGCGCGCTCGATCTTGT
30 CGAGCAGGATGAAACGAACCTGCCGGAGCAGTCAAAGATGACCACACTTCTGGCAATCGTGCA
GAAACTGCCGACGTGATGGTGGTTTCACGGGCCCGGCTCCAACTCGCCATCGCGTGGGCGAA
CATTCTGCTGCAACGGCCGGCGCCGGCCACCGGTGCCCTGAATCGTTTCGAAACGGCCCTTGG
CCGGGCCGAGCTTCCCGAGGCGACGCGAGCGGATCTGCGGGCCGAGGCAGACGTGTTGCGG
GCGGTGCGCGAGGTGTTGCGAGACCGGGTCGAGCGCGTGGATGACCTTCTCGCCGAGGCAAT
35 GTCGAGACCGGACACCCTGCCCCGCGAGTCCCCGGGACCGCCGGCAACACCGCGGCGTTGG
CCGCGATCTGCCGCTTCGAGTTCGCCGAGGTATATCCACTGCTGGAAGTGGGCCGCGCCCTACC
AGGAAATGATGGGACCGTTCGGCACCGTTTATGCGCAGTGCTTGCGCGGCATGGCGGCCAGGA

ATCGGCTCGACATTGTCGCTGCGCTACAGAACTTCCGAACGGCGTTTCGAGGTCGGCACGGCAG
TGGGGGGCCCACTCGCACGCGGCGCGGCTTTCGGGTTTCGCTGCTCGCCGAATTGCTCTACGAG
ACCGGCGATCTGGCCGGGGCTGGTCGTCTCATGGACGAGAGCTATCTGCTGGGTTCCGAGGG
GGGTGCAGTGGAACCTGGCCGCCAGGTACGTGATCGGCGCGCGGGTCAAGGCGGGCCAGG
5 GGGATCATGAGGGTGCGBCTGATCGCCTGTCCACCGAGGCGATACTGCCGTCCAGCTGGGG
CTGCCGCGCCTGGCTGCCCGAATCAACAACGAGCGGATCCGGCTGGGCATCGCGCTACCTGC
GGCGGTGGCCGCCGATTTGCTGGCACCCCGCACCATCCCCCGCGACAATGGAATCGCCACCAT
GACAGCCGAACCTCGACGAGGACTCCGCGGTGCGCCTGTTGTCCGCCGGCGACTCCGCCGATC
GTGACCAAGCCTGCCAACGGGCGCGGTGCTCTCGCCGCCGCCATCGACGGTACGCGCAGACCG
10 CTGGCGGCGCTGCAGGCGCAAATACTTCATATCGAAACGCTTGCCGCCACCGGACGGGAATCC
GATGCGCGAAACGAACCTGGCGCGGTAGCCACGAAGTGCGCCGAACCTCGGGCTGTCACGTCT
GCTGGTCGATGCGGGACTGGCCTAA

>Rv3106 fprA adrenodoxin and NADPH ferredoxin reductase TB.seq 3474004:3475371
15 MW:49342 >emb|AL123456|MTBH37RV:3474004-3475374, fprA SEQ ID NO:120
ATGCGTCCCTATTACATCGCCATCGTGGGCTCCGGGCGGTCGGCGTTCTTCGCCGCGGCATCC
TTGCTGAAGGCCGCGACACGACCGAGGACCTCGACATGGCCGTCGACATGCTGGAGATGTTG
CCGACTCCCTGGGGGCTGGTGCCTCCGGGGTCGCGCCGGATCACCCCAAGATCAAGTCGAT
CAGCAAGCAATTCGAAAAGACGGCCGAGGACCCCGCTTCCGCTTCTTCGGCAATGTGGTCGT
20 CGGCGAACACGTCCAGCCCGGCGAGCTCTCCGAGCGCTACGACGCCGTGATCTACGCCGTGCG
GCGCGCAGTCCGATCGCATGTTGAACATCCCCGGTGAGGACCTGCCGGGCAGTATCGCCGCC
GTGATTTTCGTGGCTGGTACAACGCACATCCACACTTCGAGCAGGTATCACCCGATCTGTGCG
GCGCCCGGGCCGTAGTTATCGGCAATGGAAACGTGCGGCTAGACGTGGCACGGATTCTGCTCA
CCGATCCCGACGTGTTGGCACGCACCGATATCGCCGATCACGCTTTGGAATCGCTACGCCAC
25 GCGGTATCCAGGAGGTGGTGATCGTCGGGCGCCGAGGTCCGCTGCAGGCCGCGTTACCACG
TTGGAGTTGCGCGAGCTGGCCGACCTCGACGGGGTTGACGTGGTGATCGATCCGGCGGAGCT
GGACGGCATTACCGACGAGGACGCGGCCGCGGTGGGCAAGGTCTGCAAGCAGAACATCAAGG
TGCTGCGTGGCTATGCGGACCGCGAACCCCGCCCGGGACACCGCCGCATGGTGTTCCGGTTCT
TGACCTCTCCGATCGAGATCAAGGGCAAGCGCAAAGTGAGCGGATCGTGCTGGGCCGCAACG
30 AGCTGGTCTCCGACGGCAGCGGGCGAGTGCGGGCCAAGGACACCGGCGAGCGCGAGGAGCT
GCCAGCTCAGCTGGTCGTGCGGTGCGTACCGCGGGGTGCCACGCCCGGGCTGCCGT
TCGACGACCAGAGCGGGACCATCCCCAACGTGCGCGGCCGAATCAACGGCAGCCCCAACGAAT
ACGTCGTGGGTGGATCAAGCGCGGGCCGACCGGGGTGATCGGGACCAACAAGAAGGACGCC
CAAGACACCGTCGACACCTTGATCAAGAATCTTGCCAACGCCAAGGAGGGCGCCGAGTGCAAG
35 AGCTTTCCGGAAGATCATGCCGACCAGGTGGCCGACTGGCTAGCAGCACGCCAGCCGAAGCTG
GTCACGTGCGCCCACTGGCAGGTGATCGACGCTTTCGAGCGGGCCGCGCGGAGCCGCACGG
GCGTCCCCGGGTCAAGTTGGCCAGCCTGGCCGAGCTGTTGCGGATTGGGCTCGGCTGA

>Rv3235 - TB.seq 3611296:3611934 MW:22659 >emb|AL123456|MTBH37RV:3611296-3611937,
Rv3235 SEQ ID NO:121

5 ATGATGGCCAGCAACCAAACCGCTGCGCAACACTCGTCTGCCACTCTCCAGCAGGCTCCTCGTT
CGATCGATGATGCTGGAGGGTGGCCCTTGACCATCAGTCCTATCGCGAACTCACCGGGCGACA
CCTTCGCCGTCACACCCGTCGTCGAGTACGAGCCGCCGCCGCGAAACATCCCGCCGTGCGGG
CAATCATCGCACGCAGCCCGGCGGGCCGCACACCCCGCAGCTAGCTCGCCGACAACCAATCAGG
CCGAGCGGCCGGGCACCGGCAGCGGTACCTCCACGGCCAAGTCACCGCGGCTGCGTCAAGC
10 GGGGACCTTCGCCGATGCCGCGCTACGCCGAGTGTGGAGGTCATCGACCGCCGCCGCCCGG
TGGGCCAGCTGCGCCCCCTGCTGGCACCCGGCCTCGTCGACTCCGTGCTCGCGGTGAGCCGC
ACGGCGGCCCGGACACCAACAAGGCGCGGCCATGCTGCGCCGCATCCGGCTGACACCGGCCGG
ACCCGACACCGCGGACACCGCCGCCGAGGTCTTCGGCACCTACAGTCGCGGGGACCGGATCC
ATGCGATCGCCTGCCGGGTGGAACAACGGCCCGCCGGTAACGAAACCCGATGGCTGATGGTC
GCCCTGCACATCGGGTGA

15

>Rv3255c manA mannose-6-phosphate isomerase TB.seq 3635040:3636263 MW:43340

>emb|AL123456|MTBH37RV:c3636263-3635037, manA SEQ ID NO:122

20 GTGGAAGTGTACGTGGCGCGTTACGCACCTACGCTTGGGGATCGCGCACCGCTATCGCCGAA
TTCACCGGGCGTCCGGTGCCGGCCGCTCACCCCGAGGCCGAACTATGGTTCCGGTGACACCC
GGGTGATCCGGCTTGGCTGCAGACGCCGCATGGCCAAACCTCGTTGCTCGAAGCGTTGGTCGC
GGATCCGGAGGGGCGAGCTCGGCTCCGCGTCGCGCGCGGATTGCGCGATGTGTTGCCGTTCT
TGGTCAAGGTGTTGGCGGCCGACGAGCCACTATCGTTGCAGGCCCATCCGAGCGCCGAGCAG
GCGGTTGAGGGCTACCTGCCGGAAGAGCGAATGGGCATTCCGGTGTCCTCACCCGTCCGCAAC
TACCGCGACACCAGTCACAAGCCAGAGTTATTGGTGGCGCTGCAGCCGTTGAGGGCGCTGGCC
25 GGATTCCGGGAGGCGGGCTCGCACCAACGAGCTGCTGCGGGCGCTGGCCGTATCCGACCTCGA
CCC GTTCATCGACTTGCTGAGCGAGGGGTCCGATGCCGATGGTTTGCGTGCGCTGTTACCCAC
CTGGATTACCGCACCCCGAGCCCGACATCGACGTGCTGGTGCCCTGCCGTGCTGGACGGCGCTAT
CCAGTACGTCAGCTCCGGCGCAACGGAATTTGGCGCCGAAGCCAAGACAGTGTGGAACTCGG
CGAACGTTATCCCGGCGACGCCGGTGTGCTGGCGGCGTTGTTGCTCAACCGCATCAGCTTGGC
30 TCCTGGGGAGGCGATCTTCCTGCCGGCCGGCAACCTGCACGCCTATGTGCGTGGTTTCGGTGT
GGAAGTGATGGCCAACTCCGACAACGTGTTACGCGGTGGACTTACCCCTAAGCACGTGATGT
GCCCCAGTTGTTGCGGGTGCTGGACTTCGCCCCACGCCGAAGGCTCGGCTGCGGCCCCCGA
TCCGGCGCGAGGGGCTGGGGCTGGTCTTTGAGACGCCACCGATGAGTTCGCGGCCACGCTA
CTGGTGCTCGACGGCGATCACCTCGGCCACGAGGTGACGCGTCGTCCGGCCATGACGGTCC
35 ACAGATCTTGTTATGCACCGAGGGTTCCGGCGACGGTGACGCGGAAGTGCGGGTTCGCTCACGCT
ACAGCGCGGCACGGCCGCCTGGGTGGCGGCCGACGACGGCCCGATCCGGCTGACCGCCGGC
CAACCCGCCAAGCTGTTCAAGGCGACCGTCGGGTTGTGA

>Rv3264c mliA2 glucose-1-phosphate thymidyltransferase TB.seq 3644897:3645973 MW:37840

>emb|AL123456|MTBH37RV:c3645973-3644894, mliA2 SEQ ID NO:123

TTGGCAACTCACCAAGTCGATGCGGTGGTCTGGTGGTGGCAAGGGTACCCGACTGCGGCCG
5 TTGACGCTGTCGGCGCCCAAGCCAATGCTGCCTACCGCCGGACTGCCGTTCTCACCCATCTG
CTGTCGCGGATCGCCGCAGCGGGCATCGAGCACGTGATCCTGGGTACGTCCTACAAACCCGCA
GTCTTCGAAGCGGAGTTCGGCGACGGGTCCGCACTGGGCCTACAGATCGAATACGTGACCGAG
GAGCATCCCTTGGGGACTGGCGGCGGCATCGCCAACGTTGCCGGCAAGCTGCGCAACGACAC
CGCGATGGTGTTTAACGGCGATGTGCTCTCGGGCGCGGATCTGGCCCAACTGCTGGACTTCCA
10 CCGAAGCAATCGAGCCGATGTCACGCTGCAACTGGTGCGGGTGGGCGACCCGCGGGCATTG
GCTGCGTACCCACCGACGAGGAGGACCGCGTAGTCGCCTTTCTGGAGAAGACGGAGGATCCG
CCGACCGACAGATCAATGCCGGCTGCTATGTCTTCGAACGCAACGTCATCGACCGGATTCCGC
AGGGCCCGGAGGTTTTCGGTGGAACGCGAGGTGTTCCCGGCCTTGCTCGCCGACGCGGACTGC
AAGATCTACGGCTATGTGATGCCAGCTATTGGCGGGACATGGGCACACCGGAAGACTTCGTTG
15 GCGGATCGGCGGATCTGGTGCGCGGCATCGCCCCGTCTCCGGCCTTGCGTGGTCACCGCGGT
GAGCAGTTGGTGACGACGGTGCGGCGGTATCTCCCGGTGCGTTGCTGATTGGCGGCACCGTC
GTGGGGCGTGGTGCCGAAATCGGCCCCGGCACCAGATTGGACGGCGCGGTGATCTTCGATGG
TGTCGGGTGGAGGCCGGGTGCGTGATCGAGCGTTGATCATCGGCTTCGGTGCTCGCATCGG
ACCGCGGGCGTTGATCCGCGACGGTGATCGGTGACGGGGCCGACATCGGCGCGCGCTGCG
20 AGTTGTTAAGTGGTGCCCGGTATGGCCCGGTGTCTTTCTTCCCGACGGCGGGATCCGTTACTC
GTCCGACGTTTGA

>Rv3368c - TB.seq 3780334:3780975 MW:23734 >emb|AL123456|MTBH37RV:c3780975-3780331,
Rv3368c SEQ ID NO:124

25 ATGACCCTCAACCTGTCCGTGACGAGGTCTGACCACTACCCGCTCGGTGCGCAAGCGTCTC
GATTTGACAAGCCGGTGCCACGCGACGTGCTGATGGAATGCCTCGAGCTGGCGCTGCAGGCG
CCCACCGGTTCCAATTCCCAAGGCTGGCAGTGGGTGTTGTCGAGGACGCCGCCAAGAAAAAG
GCGATCGCCGACGTCTACCTGGCCAACGCCCGGGGCTACCTCAGCGGGCCGGCGCCCGAGTA
CCCCGACGGCGACACCCGCGCGGAGCGGATGGGGCGGGTCCGCGATTGCGCGACCTATCTCG
30 CCGAACACATGCACCGGGCGCCGGTGCTGCTGATCCCTGCCTGAAAGGCCGGGAAGACGAG
TCGGCGGTGGGTGGCGTGCTGTTTTGGGCCTCACTGTTCCCGCGGTGTGGAGCTTCTGCCTG
GCGCTGCGCTCCCGCGGGGCTGGGTTGCTGCTGGACGACGCTGCACCTGCTCGACAACGGCGA
GCACAAGGTGGCCGACGTGCTCGGCATTCCCTACGACGAATACAGCCAAGGCCGGGCTGCTTCC
GATCGCCTACACACAAGGCATCGACTTCCGGCCGGCCAAGCGGCTGCCGGCCGAGAGCGTGA
35 CGCACTGGAACGGCTGGTAA

>Rv3382c lytB1 TB.seq 3796447:3797433 MW:34667 >emb|AL123456|MTBH37RV:c3797433-3796444, lytB SEQ ID NO:125

ATGGCTGAGGTGTTTCGTGGGACCGGTCGCACAGGGATACGCTTCGGGTGAAGTCACGGTGCTG
TTGGCGTCGCCGCGGTGCTTTTGCGCCGGTGTAGAGCGTGCTATCGAGACGGTCAAGCGAGTG
5 CTTGACGTGGCCGAAGGCCCGGTGTATGTGCGCAAGCAAATCGTGACAACACTGTTGTGGTT
GCCGAGTTGCGGGACCGGGGAGCAGTGTTCGTGAGGATCTCGACGAGATTCCCGATCCGCC
GCCGCCGGGGGCGGTGCTGGTGTCTCCGCGCATGGGGTTTCCCCGGCGGTGCGCGCGGGC
GCTGATGAGCGGGGACTGCAGGTCGTGACGCGACCTGCCCACTGGTGGCGAAAGTCCACGC
TGAAGCCGCACGGTTTCCGCGCGCGGTGACACGGTGGTCTTCATCGGGCACGCCGGACATG
10 AGGAGACCGAAGGCACGCTTGGCGTCGCTCCGCGGTCAACATTATTGGTGCAGACACCCGCTG
ATGTGGCAGCGTTGAACCTGCCCGAGGGTACCCAGCTATCGTATCTGACCCAGACAACCCTGG
CACTTGATGAACTGCCGATGTCATTGATGCGCTGCGCGCGAGGTTTCCGACGTTGGGCCAACC
CCCCTCTGAAGACATCTGCTATGCCACCACGAACAGACAGCGTGCGCTGCAATCGATGGTCGGT
GAATGTGACGTTGTGTTGGTGATTGGCTCGTGCAATTCGTGCAATTCGCGGCGTCTGGTCGAGT
15 TGGCGCAGCGAAGTGGGACGCCGGCCTACTTGATTGACGGGCCTGATGACATTGAGCCCGAAT
GGCTGTGTCGGTCTCGACGATCGGTGTCACCGCGGGAGCCTCCGCGCCGCCACGACTGGTG
GGGCAGGTGATTGATGCACTTCGCGGATACGCCTCGATACCGTGGTGGAACGCTCGATAGCG
ACCGAGACGGTGCGATTGCGCCTTCCCAAACAGGTTGCGCGCGCAATGA

20 >Rv3418c groES 10 kD chaperone TB.seq 3836985:3837284 MW:10773

>emb|AL123456|MTBH37RV:c3837284-3836982, groES SEQ ID NO:126

GTGGCGAAGGTGAACATCAAGCCACTCGAGGACAAGATTCTCGTGCAAGGCCAACGAGGCCGAG
ACCACGACCGCGTCCGGTCTGGTCATTCTGACACCGCCAAGGAGAAGCCGCAGGAGGGCAC
CGTCGTTGCCGTCGGCCCTGGCCGGTGGGACGAGGACGGCGAGAAGCGGATCCCGCTGGACG
25 TTGCGGAGGGTGACACCGTCATCTACAGCAAGTACGGCGGCACCGAGATCAAGTACAACGGCG
AGGAATACCTGATCCTGTGCGGCACGCGACGTGCTGGCCGTGCTTTCCAAGTAG

>Rv3423c alr TB.seq 3840193:3841416 MW:43357

>emb|AL123456|MTBH37RV:c3841416-3840190, alr SEQ ID NO:127

GTGAAACGGTTCTGGGAGAATGTCGGAAAGCCAAACGACACGACAGATGGGCGGGGCACGACT
TCGTTGGCCATGACACCGATATCCCAGACACCTGGCCTCCTCGCCGAGGCCATGGTGGATCTG
GGCGCTATTGAACACAACGTGCGGGTGCTGCGTGAGCACGCCGGCCACGCGCAGCTGATGGC
GGTGGTCAAGGCCGACGGCTACGGTACGGTGCTACGCGCGTCGCCAAACCGCCCTGGGAG
CCGGTGCGGCCGAACCTCGCGCTCGCCACCGTCGACGAGGCGCTAGCGCTGCGCGCTGATGGC
35 ATTACCGCACCGGTGCTGGCCTGGCTGCATCCGCCCGGCATCGACTTCGGGCCCGCGCTGCTG
GCCGACGTGCAGGTGCGGGTGTCTCGCTGCGCCAACCTCGACGAACTGTTGCACGCGGTGCG
CCGGACCGGCCGGACGGCGACGGTGACCGTCAAGGTGGATACCGGGCTGAACCGCAATGGCG

TGGGACCGGCACAATTCCCGGCCATGCTGACCGCGTTACGCCAAGCCATGGCCGAGGACGCC
GTCCGGCTGCGGGGGCTGATGTCGCATATGGTTTACGCCGACAAGCCTGACGATTCCATCAAC
GATGTTCAAGGCCCAACGGTTTACCGCCTTTCTGGCGCAGGCCCCGGAACAAGGGGTGCGGTTC
GAGGTGGCGCATCTATCGAACTCATCAGCAACTATGGCGCGCCCCGACCTGACGTTTCGACCTG
5 GTGCGGGCCGGGCATCGCGGTGTATGGGCTAAGCCCGGTACCCGCCCTCGGTGACATGGGGCT
GGTGCCGGCGATGACCGTGAAATGTGCTGTTGCGCTGGTGAAATCGATTCTGTGCGGGGGAGGG
CGTGTCGTATGGGCACACATGGATCGCGCCACGCGACACCAATCTGGCGCTGCTGCCGATCGG
TTACGCAGACGGCGTGTTCCGGTCGCTGGGCGGGCGGCTGGAGGTGCTGATCAACGGCAGAC
GATGCCCCGGTGTTGGGGCGGATCTGCATGGACCAGTTCATGGTCGACCTGGGCCCCGGGCCG
10 CTTGATGTGGCCGAAGGCGACGAGGCGATTGTTTGGTTCGGGCCGGGCATCCGGGGTGAGCCAC
GGCTCAGGACTGGGCCGATCTTGTCGGCACCATCCACTACGAAGTGGTCACCAGCCCCGCGAGG
ACGTATCACCAGGACCTATCGCGAGGCTGAAAACCGTTGA

>Rv3490 otsA [alpha]-trehalose-phosphate synthase TB.seq 3908232:3909731 MW:55864

15 >emb|AL123456|MTBH37RV:3908232-3909734, otsA SEQ ID NO:128

ATGGCTCCCTCGGGAGGCCAGGAGGCGCAGATTTGCGATTGGAGACCTTCGGGGACTCTGAC
TTCGTGGTGGTAGCCAATCGACTGCCCGTCGATCTGGAGCGTCTTCCCGACGGCAGCACAACC
TGGAACGCAGCCCCGGAGGCTTGGTACCGCCTTGAGCCGGTGCTGCGGCGTCGGCGCGG
GGCCTGGGTGGCTGGCCCCGGCGTTAACGACGACGGGGCCGAACCCGACCTCCACGTGCTGG
20 ACGGCCCCATCATCCAAGACGAGCTGGAACCTCATCCGGTACGGCTGAGCACCACGGACATAG
CTCAGTACTACGAGGGATTCTCCAACGCCACACTGTGGCCGCTGTACCACGACGTCATCGTCAA
GCCGCTCTACCACCGCAATGGTGGGATCGCTACGTCGACGTCAACCAGCGCTTTGCCGAGGC
CGCGTCGCGCGCCGCGCCACGGCGCAACCGTGTGGGTACAGGACTACCAGCTGCAGCTGG
TACCGAAGATGCTGCGCATGCTGCGGCCCGATCTGACCATCGGTTTCTTTTGCACATCCCGTT
25 CCGGCCGGTAGAGCTGTTTATGCAGATGCCGTGGCGCACCGAGATCATCCAGGGCCTACTGGG
CGCCGACCTGGTGGGCTTCCATCTTCCGGGCGGTGCCCAGAATTCCTGATCCTGTCCCGGCG
TCTGGTGGCACCACACTTCCCGCGGAACCGTCGGTGTGCGGTGCGGTTCCGTGCGGCGG
TGCTCGGGTCCCGCACCATACGAGTTGGCGCCTTTCCTATCTCGGTTGACTCCGGCGCGCTCG
ACCACGCTGCCCGCGACCGCAACATCAGGCGCGGGCCCCGCGAGATTGCAACCGAACTGGGA
30 AATCCGCGCAAGATCCTGCTCGGTGTTGACCGGCTCGACTACACCAAGGGCATCGACGTACGG
CTGAAGGCCTTTTCCGAGCTGCTGGCCGAGGGCCGCGTCAAACGCGACGACACCGTCGTGGTC
CAGCTGGCTACCCCGAGCCGCGAGCGGGTGGAGAGCTACCAGACGCTGCGCAACGACATCGA
ACGCCAGGTCGGCCACATTAACGGCGAGTACGGTGAGGTTGGCCATCCGGTAGTGATTACCT
GCATCGACCGGCTCCGCGCGACGAGCTTATCGCTTTCCTGCGGCCAGCGACGTCATGCTGGT
35 CACCCCACTACGCGACGGGATGAACCTGGTGGCCAAGGAGTACGTCGCTTGCCGCGAGCGATCT
TGGCGGTGCCCTGGTGCTCAGCGAATTCACCGGGGCGCGAGCCGAACCTCCGGCACGCATACCT
GGTCAACCCGCACGACCTGGAAGGCGTCAAGGACGGGATAGAGGAAGCGCTCAACCAGACGG

AGGAGGCGGGCCGGCGGCGAATGCGGTGCTGCGACGCCAAGTGCTCGCCACGACGTGGA
CCGCTGGGCACAGTCGTTTCTCGACGCTCTCGCCGGGGCACACCCGAGGGGCCAAGGCTAA

>Rv3598c lysS lysyl-tRNA synthase TB.seq 4041423:4042937 MW:55678

5 >emb|AL123456|MTBH37RV:c4042937-4041420, lysS SEQ ID NO:129
GTGAGTGCCGCTGACACAGCAGAAGACCTTCCTGAGCAGTTCGGATTGCGCGGGACAAGCGC
GCTCGCTTGCTGGCCCAGGGGCGCGATCCCTATCCCGTCGCGGTGCCGCGCACTCACACGTTG
GCCGAGTTTCGCGCCGCCACCCTGACTTGCCGATCGATACCGCGACCGAAGACATCGTCGGC
GTGCGGGGCCGAGTGATCTTTGCGCGCAACTCGGAAAGCTATGCTTTGCGACACTTCAGGAC
10 GCGATGGTACCCAGCTGCAAGTGATGATCAGCCTCGACAAGGTCGCGCCAGGCTGCTCTCGAC
GCATGAAAGCCGATGTGACCTGGGCGACATCGTCTACGTGCATGGCGCGGTGATCAGTTCG
CGCCGCGGCGAGCTGTCCGTCTGGCGGATTGCTGGCGGATCGCCGCCAAGTCGCTGCGGCC
GCTTCCCGTCGCGCACAAAGAGATGAGTGAAGAGTCGCGGGTTCGTCAGCGCTATGTTGACCT
CATAGTTCGACCGGAAGCGCGCGCGGTGGCTCGACTACGGATCGCCGTCGTCCGCGCGATCC
15 GGACGGCGCTTCAACGTCGTGGGTTCCTGGAAGTCGAGACGCCCGTCTTGACACGTTAGCCG
GTGGTGCGGCGGCCCGTCCGTTCCGCACTCATTCCAATGCCCTAGACATCGATCTGTACCTGCG
GATCGCGCCGGAAGTGTTCCTCAAGCGCTGCATCGTGGGTGGTTTCGACAAGGTCTTCGAAGTT
AATCGAGTGTTCCGAAACGAAGGAGCCGATTCCACGCATTCTCCGGAATTCTCCATGCTGGAGA
CCTACCAGACCTACGGAACCTATGACGATTCGGCAGTCGTCACCCGGGAGCTTATTCAAGAGGT
20 GGCCGATGAGGCGATCGGAACCAGACAAGTCCGTTGCCCGACGGCAGTGTCTATGACATCGA
CGGAGAATGGGCGACTATACAAATGTACCCGTCGCTGTCTGTGGCGCTCGGTGAAGAGATCAC
ACCGCAGACGACGGTCGATCGCTTACGTGGGATCGCCGATAGCCTTGGCCTGGAGAAAGACCC
AGCGATTATGACAACCGTGGCTTCGGCCACGGCAAACCTCATCGAGGAACTCTGGGAGCGCAC
AGTGGGCAAGAGCTTGAGCGCACCCACATTTGTCAAGGATTTTCCGGTTCAGACAACGCCTTTG
25 ACCCGTCAGCACCGCAGTATCCCCGGCGTAACCGAGAAGTGGGACCTCTATCTGCGCGGAATC
GAACTTGCCACCGGCTACTCGGAATTAAGCGACCCGGTAGTCCAGCGGGAGAGATTCGCCGAC
CAGGCCCGTGCCGCGGCCGCTGGCGATGACGAAGCGATGGTGCTTGACGAGGATTTTCTGGCC
GCTCTGGAGTACGGCATGCCACCGTGCACCGGAACCGGAATGGGTATCGATCGGTTGTTGATG
TCTTTGACTGGGTGTCAATTAGGGAGACAGTTTGTTCGCGATTGTTGACACACTCCAAGT
30 A

>Rv3600c - similar to Bacillus subtilis protein YacB TB.seq 4043041:4043856 MW:29274

>emb|AL123456|MTBH37RV:c4043856-4043038, Rv3600c SEQ ID NO:130

GTGCTGCTGGCGATTGACGTCCGCAACACCCACACCGTTGTGGGCCTGCTGTCCGGAATGAAA
35 GAGCACGCAAAGGTCGTGCAGCAGTGGCGGATACGCACCGAATCCGAAGTCACCGCCGACGAA
CTGGCACTGACGATCGACGGGCTGATCGGCGAGGATTCCGAGCGGCTACCGGTACCGCCGC
CTTGCCACGGTCCCGTCCGTGCTGCACGAGGTGCGGATAATGCTCGACCACTACTGGCCGTC

GGTGCCGCACGTGCTGATCGAGCCCGGAGTACGCACCGGGATCCCTTTGCTCGTCGACAACCC
GAAGGAAGTGGGCGCAGACCGCATCGTGAACGTGTTTGGCCGCCTATGACCGGTTCCGGAAGGC
CGCCATCGTCGTTGACTTTGGATCCTCGATCTGTGTTGATGTTGTATCGGCCAAGGGTGAATTC
TTGGCGGCGCCATCGCGCCCGGGGTGCAGGTGTCTCCGATGCCGCGGCGGCCCGCTCGGCG
5 GCATTGCGCCGCGTTGAACTTGCCCCGCCACGTTCCGGTGGTTGGCAAGAACACCGTCGAATGC
ATGCAAGCCGGTGCAGGTGTTCCGGCTTCGCCGGGCTGGTAGACGGGTTGGTAGGCCGCATCCG
CGAGGACGTGTCCGGTTTCTCCGTCGACCACGATGTCCGCGATCGTGGCTACCGGGCATACCGC
GCCCCTGCTGCTGCCGGAATTGCACACCGTCGACCATTACGACCAGCACCTGACCTTGACAGGG
TCTGCGGCTGGTGTTCGAGCGTAACCTCGAAGTCCAGCGCGGCGCGGCTCAAGACGGCGCGCT
10 GA

>Rv3606c folK 7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase TB.seq
4048181:4048744 MW:20732 >emb|AL123456|MTBH37RV:c4048744-4048178, folK
SEQ ID NO:131

15 ATGACGCGGGTAGTGCTCTCGGTTGGCTCCAACCTGGGTGACCGCCTGGCACGATTGCGGTGCG
GTCGCCGACGGTCTCGGCGATGCGTTGATTGCGGCTTCCCCGATATATGAGGCCGACCCCTGG
GGTGGGGTGGAGCAGGGGCGAGTTCTCAATGCGGTGCTGATCGCCGACGATCCTACCTGCGAA
CCGCGGGAGTGGCTGCGGCGGGCGCAGGAGTTGAGCGCGCTGCGGGCAGGGTGCGTGGCC
AGCGCTGGGGTCCACGAAATCTCGACGTGACCTGATCGCCTGCTACCAGACCTCGGCCACCG
20 AGGCTCTGGTGAAGTGACCGCGCGGGAGAACCACCTCACGCTGCCGCACCCACTGGCGCAT
CTGCGGGCCTTTGTGTTGATCCCGTGGATTGCCGTGACCCAACGGCGCAGCTGACGGTTGCC
GGGTGCCCGCGGCCCGTCACGCGACTGCTGGCCGAGCTGGAGCCCGCCGACCGCGACAGTGT
GCGGTTGTTTAGGCCGTGTTGATCTGAATAGCAGACACCCCGTCAGTCGGGCACCGGAAAG
CTGA

25 >Rv3607c folX may be involved in folate biosynthesis TB.seq 4048744:4049142 MW:14553
>emb|AL123456|MTBH37RV:c4049142-4048741, folX SEQ ID NO:132

30 ATGGCTGACCGAATCGAACTGCGCGGCCTGACCGTGCATGGTCGGCACGGGGTCTACGACCAC
GAGCGAGTGGCCGGGCAGCGGTTTGTTCATCGATGTCACCGTGTGGATAGACCTGGCCGAGGC
CGCCAACAGCGACGACTTGGCCGACACCTATGACTACGTGCGGCTGGCTTCGCGGGCGGCGCG
AGATCGTCGCCGACCCCCGCGGAAGCTGATCGAAACGGTCGGGGCCGAGATCGCTGATCAC
GTGATGGACGACCAGCGAGTGCATGCCGTTGAGGTGGCGGTACACAAGCCGCAGGCGCCCAT
CCGACAGCTTCGACGATGTGGCGGTGGTATCCGACGCTCACGGCGCGGCGGCGCGGTTG
GGTAGTCCCGGCGGGCGGCGCGGTATGA

35 >Rv3608c folP dihydropteroate synthase TB.seq 4049138:4049977 MW:28812
>emb|AL123456|MTBH37RV:c4049977-4049135, folP SEQ ID NO:133

GTGAGTCCGGCGCCCGTGCAGGTGATGGGGTTCTAAACGTCACGGACGACTCTTTCTCGGAC
 GGGCGGTGTTATCTCGATCTCGACGATGCGGTGAAGCACGGTCTGGCGATGGCAGCCGCAGGT
 GCGGGCATCGTCGACGTCGGTGGTGAGTCGAGCCGGCCCGGTGCCACTCGGGTTGACCCGGC
 GGTGGAGACGTCTCGTGTACATCCCGTCGTCAAAGAGCTTGACGACACAAGGCATCACCGTCAG
 5 CATCGATACCATGCGCGCGGATGTCGCTCGGGCGGCGTTGCAGAACGGTGCCCAGATGGTCAA
 CGACGTGTCGGGTGGGCGGGCCGATCCGGCGATGGGGCCGCTGTTGGCCGAGGCCGATGTG
 CCGTGGGTGTTGATGCACTGGCGGGCGGTATCGGCCGATACCCCGCATGTGCCTGTGCGCTAC
 GGCAACGTGGTGGCCGAGGTCCGTGCCGACCTGCTGGCCAGCGTCGCCGACGCGGTGGCCGC
 AGGCGTCGACCCGGCAAGGCTGGTGCTCGATCCCGGGCTTGGATTGCGCAAGACGGCGCAAC
 10 ATAATTGGGCGATCTTGATGCCCTTCCGGAAGTGGTCGCGACCGGAATCCAGTGCTGGTGG
 GTGCTTCGCGCAAGCGCTTCTCGGTGCGTTGTTGGCCGGGCCCCGACGGCGTGATGCGGCCA
 ACCGATGGGCGTGACACCGCGACGGCGGTGATTTCCGCGCTGGCCGCACTGCACGGGGCCTG
 GGGTGTGCGGGTGATGATGTGCGGGCCTCGGTGATGCCATCAAGGTGGTCAAGCGTGGAT
 GGGAGCGGAAAGGATAGAACCGCATGGCTGA

15

>Rv3609c folE GTP cyclohydrolase I TB.seq 4049977:4050582 MW:22395

>emb|AL123456|MTBH37RV:c4050582-4049974, folE SEQ ID NO:134

ATGTGCGAGCTGGATTGCGCGCAGCGCATCTGCTCGTATCCGTGTGTTGACCGCAACGTGCC
 GAGGCCCGCGGTGCGCGAATTGCTGTACGCGATCGGCGAGGATCCGGATAGGGACGGCTTGGT
 20 AGCCACCCCGTCCCGGGTTGCCCGGTCATACCGCGAAATGTTGCGCGGGCTCTACACCGACCC
 CGACTCGGTGTTGAACACCATGTTGACGAAGACACGACGAGCTGGTGTGGTCAAGGAAATC
 CCTATGTACTCCACCTGCGAACACCACCTGGTGGCGTTCCACGGTGTGGCCACGTGCGGTAC
 ATCCCGGGCGACGACGGCAGGGTGACCGGCTTGTCAAAGATCGCGCGACTGGTGCATCTGTAC
 GCCAAGCGACCTCAGGTCCAGGAGCGGCTCACCAGTCAGATCGCCGATGCCCTGATGAAAAAA
 25 CTCGATCCACGCGGGTAATCGTGGTGATCGAGGCTGAGCATCTGTGCATGGCGATGCGCGGG
 GTTCGCAAGCCCGGCTCGGTCAACCTACGTGCGCGGTGCGCGGACTGTTCAAACCAATGCC
 GCTTCTCGAGCCGAAGCGCTCGACCTCATTTTGGGAAGTGA

30

>Rv3610c ftsH inner membrane protein, chaperone TB.seq 4050601:4052880 MW:81987

>emb|AL123456|MTBH37RV:c4052880-4050598, ftsH SEQ ID NO:135

ATGAACCGGAAAAACGTGACTCGCACCATACAGCGATCGCCGTCGTGGTGCTGCTCGGCTGG
 TCGTTCTTTTACTTCAGCGACGACACCCGCGGCTACAAGCCCGTTGATACCTCGGTGGCGATAA
 CACAGATCAACGGCGACAACGTCAAGAGCGCACAGATCGACGATCGCGAGCAACAGCTGCGGC
 TGATCCTGAAGAAGGGTAACAACGAGACCGACGGTCCGAGAAGGTCATCACCAAGTACCCCA
 35 CCGGGTACGCCGTCGACCTGTTCAACGCGCTCAGCGCCAAAAACGCGAAGGTCAGCACGGTCG
 TCAACCAGGGCAGCATCCTGGGCGAGCTGCTGGTCTACGTGCTGCCGCTGCTGTTGCTGGTGG
 GGCTGTTGCTGATGTTCTCCCGCATGCAAGGCGGCGCCCGGATGGGCTTCGGGTTGGCAAGT

CACGCGCCAAGCAACTGAGCAAGGACATGCCCAAGACCACCTTCGCCGACGTCGCAGGTGTGG
ACGAGGCGGTTCGAGGAGCTCTACGAGATCAAGGACTTCCTGCAGAACCCAGCAGGTACCAAG
CGCTGGGCGCCAAGATCCCCAAAGGCGTGCTGCTCTACGGGCCGCCGGAACCGGTAAGACG
TTGCTGGCTCGTGCGGTGGCCGGCGAAGCCGGAGTGCCGTTCTTCACCATCTCCGGCTCCGAC
5 TTCGTCGAAATGTTTCGTCGGCGTCGGCGCATCCCGTGTCAGAGACCTGTTTCGAGCAGGCCAAG
CAGAACAGCCCGTGCATCATCTTCGTCGACGAGATCGACGCCGTCGGCCGACAAAGAGGCGCC
GGGCTGGGCGGCGGTACGACGAGCGTGAGCAGACCCTCAACCAGTTGCTAGTCGAAATGGA
CGGTTTTGGCGATCGCGCCGGCGTCATCCTGATCGCGGCCACCAACCGGCCCGACATCCTGGA
CCCCGGCGCTGTTGCGGCCGGGCCGCTTCGACCGCCAGATCCCGGTATCCAACCCCGATCTGG
10 CGGGTCGGCGGGCGGTGCTGCGCGTGCACTCCAAGGGCAAGCCGATGGCCGCGGACGCCGA
CCTCGACGGACTGGCCAAGCGGACCGTCGGCATGACCGGAGCCGACCTGGCCAACGTCATCA
ACGAGGCGGCGCTGCTGACCGCCCGGGAGAACGGCACCGTCATCACCGGTCCCGCCCTCGAG
GAAGCGGTGGACCGGGTGATCGGCGGCCCGCGCCGCAAAGGCCGGATCATCAGCGAGCAGGA
GAAGAAGATCACCGCCTATCACGAGGGCGGGCACACCCTGGCCGCTTGGGCGATGCCCGATAT
15 CGAGCCGATTTATAAGGTGACGATCCTGGCGCGCGGGCGTACCGGCGGGCACGCGGTGGCGG
TGCCGGAAGAAGACAAGGGCCTGCGGACCCGCTCGGAAATGATCGCGCAACTGGTGTTCCGGA
TGGGTGGGCGCGCCGCCGAAGAACTGGTGTTCGTGAGCCGACCACCGGCGCGGTGTCCGAC
ATCGAGCAGGCCACCAAGATAGCGCGCTCAATGGTCACCGAATTTGGAATGAGCTCCAAGCTG
GGCGCGGTCAAATACGGCTCCGAACACGGCGACCCGTTCTCGGACGTACCATGGGCACCCAG
20 CCGGACTACTCCCACGAGGTGCCCCGCGAGATCGACGAAGAGGTCCGCAAGCTTATCGAGGCG
GCGCATACCGAAGCGTGGGAAATCCTGACCGAATACCGCGACGTGCTGGACACTTTGGCCGGC
GAGCTGCTGAAAAGGAGACCCTGCACCGACCCGAGCTGGAAAGCATCTTCGCTGACGTGAA
AAGCGGCCGCGGCTCACCATGTTTCGACGACTTCGGTGCGCGGATCCCGTCGGACAAACCGCCC
ATCAAGACACCCGGCGAGCTCGCGATCGAACGCGGCGAACCTTGGCCCCAGCCGGTCCCCGA
25 GCCGGCGTTCAAGGCGGCGATTGCGCAGGCTACCCAAGCCGCTGAGGCCGCCCGGTCCGACG
CCGGCCAAACCGGGCACGGCGCCAACGGTTCCGCCGCGGCGACCCACCGGTCCGGTGACCGC
CAGTACGGCTCCACCCAGCCTGACTACGGTGCCCCGGCGGGCTGGCATGCGCCGGGATGGCC
CCCAAGGTCATCTCATCGGCCAGCTATAGCGGTGAACCGGCACCGACGTATCCGGGTGAGCC
CTACCCGACCGGTCAAGCCGATCCGGTTCCGATGAGTCCTCGGCGGAGCAGGATGACGAGGT
30 CAGTCGGACCAAGCCGGGCCACGGCTGA

>Rv3671c - TB.seq 4112322:4113512 MW:40722 >emb|AL123456|MTBH37RV:c4113512-4112319,
Rv3671c SEQ ID NO:136

ATGACCCCGTCGCAGTGGCTGGATATCGCCGTCTTGGCGGTGCGATTTATTGCAGCCATCTCCG
35 GCTGGCGTGCCGGTGCGCTGGGCTCAATGCTGTCGTTTGGCGGGTGCTGCTGGGCGCGACA
GCCGGCGTGCTGCTGGCGCCGCGATATCGTCAGTCAAATCAGCGCTCCGCGGGCCAACTGTTT
GCCGCGCTGTTCTGATCCTGGCACTGGTCGTAGTCGGCGAGGTGCTGGTGTGGTGCTGGGC

CGCGCCGTCCGCGGGGCGATCCGTAACCGGCCGATCOGGTTGATCGACTCGGTCAATTGGGGTA
GGGGTGCACTGGTCGTGGTGCTACCGCGGCGTGGTTGTTGGCGATGCCGCTGACACAGTC
GAAAGAGCAGCCCGAGCTGGCTGCCGCGGTGAAGGGTTCGCGGGTGCTCGCCCCGGGTCAACG
AGGCGGCACCCACCTGGCTGAAGACGGTGCCCAAGCGGCTGTGGCCCTGCTGAACACCTCC
5 GGCCTGCCCGCGGTTTTGGAGCCGTTAGCCGCACGCCGGTCATTCCAGTGGCCTCACCCGAC
CCAGCGCTGGTCAACAATCCGGTGGTGGCGGCCACCGAGCCAAGTGTGTCAAAATCCGCAGC
CTGGCACCCAGATGCCAGAAAGTGTGGAGGGCACCGGCTTCGTGATCTACCCGATCGGGTG
ATGACCAACGCGCACGTGGTGGCCGATCCAACAACGTACGGTGTATGCCGGCGACAAGCCC
TTCGAGGCCACGGTGGTGTCTACGACCCGTCGGTCGACGTAGCGATCCTGGCCGTTCCGCAC
10 TTGCCCGCCCGCCGCTGGTCTTCGCTGCGGAGCCGGCGAAAACCGGTGCCGACGTTGTGGT
GCTGGGTTATCCCGGCGGCGGCAATTTCACTGCCACACCCGCCAGGATTGCGGAGGCCATCAG
ACTCAGTGGCCCCGATATTTACGGGGACCCGGAGCCGGTTACCCGCGACGTGTACACCATCAG
AGCCGATGTGGAGCAAGGTGATTCCGGGTGGGCCCTGATCGACCTCAACGGTCAGGTGCTCGG
TGTGGTGTTCGGCGCAGCCATCGACGACGCCGAACTGGGTTTGTGCTGACGGCCGGCGAGGT
15 GGCGGGGCGAGCTTGCCAAAATCGGTGCTACCCAACCGGTGCGCACCGGGGCCTGCGTCAGCT
GA

>Rv3682 ponA2 TB.seq 4121913:4124342 MW:84637

>emb|AL123456|MTBH37RV:4121913-4124345, ponA' SEQ ID NO:137

20 ATGCCCGAGCGCCTCCCGGCCGCGATCACCGTTCTGAAGCTGGCTGGGTGCTGTCTGTTGGCC
AGTGTGCTCGCCACTGCGCTGACGTTCCCGTTTCGACGGCGGGCTAGGGCTGATGTCCAATCGT
GCCTCTGAGGTGCTTGCCAACGGCTCGGCCAGCTGCTCGAGGGGCAAGTGCCTGCGGTATCG
ACGATGGTCGACGCGAAGGGCAACACGATCGCGTGGCTGTACTCGCAGCGCCGGTTTCGAGGT
GCCCTCGGACAAGATCGCCAACACGATGAAGCTGGCGATCGTCTCGATTGAAGATAAGCGTTTC
25 GCCGACCACAGCGGCGTGGACTGGAAGGGCACCCGTGACCGGCCTGGCGGGCTACGCGTCCG
GCGACCTCGACACGCGCGGCGGCTCGACGCTCGAACAACAGTACGTGAAGAACTACCAACTGC
TGGTGACAGCCCAAACCGATGCCGAGAAGCGAGCGGCCGTGAAACCACTCCGGCCCCGAAG
CTTCGCGAGATCCGGATGGCACTCACGCTGGACAAGACCTTCACAAAATCTGAAATCCTGACCC
GATACTTGAACCTGGTCTCGTTCCGGCAATAACTCGTTCCGGCGTGACGACGCGGCGCAAACGTA
30 CTTCCGCATCAACGCGTCCGACCTGAATTGGCAGCAAGCGGCGCTGCTGGCCGGCATGGTGCA
ATCGACCAGCACGCTCAACCCGTACACCAACCCCGACGGCGCGCTGGCCCCGGCGGAACGTGG
TCCTCGACACCATGATCGAGAACCTTCCCGGGGAGGCGGAGGCGTTGCGTGCCGCCAAGGCC
GAGCCGCTGGGGTACTGCCGAGCCCAATGAGTTGCCGCGCGGCTGCATCGCGGCCGGCGA
CCGCGCATTCTTCTGCGACTACGTCCAGGAGTACCTGTCTCGGGCCGGGATCAGCAAGGAGCA
35 GGTGCGCCACGGGCGGGTACCTGATCCGCAACCACCTGGACCCAGAGGTGCAGGCACCGGTCA
AGGCCGCCATCGACAAGTACGCCAGCCGAACCTGGCCGGTATTTCCAGCGTGATGAGCGTGA
TCAAACCGGGTAAGGATGCGCACAAGGTGTTGGCCATGGCCAGTAACCGCAAATACGGGCTGG

ATCTAGAAGCCGGCGAAACCATGCGGGCCGAGCCATTCTCCCTGGTTGGCGACGGCGCCGGGT
CTATCTTCAAGATCTTCAACCACGGCCGCTGCTCTGGACATGGGCATGGGTATTAACGCCCAACT
CGACGTGCCGCCCCGATTCCAGGCCAAAGGTCTGGGAAGTGGCGGGGCAAAGGGGTGCCCA
AAGAGACCTGGTGTGTGGTGAACGCCGGCAACTACCGCGGCTCGATGAATGTCACCGACGCGC
5 TGGCAACCTCGCCAAACACCGCGTTCCGCAAGCTGATCTCGCAGGTCGGGGTGGGGCGTGCG
GTGATATGGCCATCAAACCTCGGGCTGAGGTCTTATGCGAATCCCGGCACCGCACGCGACTAC
AACCCCGACAGCAATGAGAGCTTGGCTGACTTCGTCAAACGACAGAACCTGGGTTCGTTACCC
TCGGCCCCATCGAGTTAAACGCGCTGGAGCTGTCCAACGTGGCGGCCACGTTGGCATCCGGCG
GCGTGTGGTGGCCCCCAACCAATCGACCAGCTCATCGACCGCAACGGCAACGAAGTCGCGG
10 TCACCACCGAGACGTGCGACCAGGTGGTGGCCGAGGGCTGGCGAACACCCCTCGCCAACGCG
ATGAGCAAGGACGCCGTGGGCAGCGGCACGGCGGCCGTTTCGGCCGGCGCGGGGGCTGGG
ATCTGCCGATGTCCGGCAAAACCGGCACCAACGAGGCGCACCGGTCCGGCCGGCTTCGTGGGC
TTCACCAACCGCTACGCGGCGGCGAACTACATCTACGACGACTCCAGCTCGCCGACAGATCTGT
GTTCCGGCCCCGCTGCGCCATTGCGGCAGCGGCGACTTGTACGGCGGCAACGAGCCATCCCGC
15 ACCTGGTTCCCGCGATGAAGCCGATCGCCAACAACCTTCGGCGAAGTGCAGCTACCACCGACC
GATCCACGCTATGTCGACGGCGCACCAAGGCTCACGGGTACCAAGCGTGGCCGGTCTGGATGTC
GACGCCGCACGCCAGCGCCTCAAGGACGCGGGCTTCCAGGTCCCGACCAAACCAACTCGGT
CAACAGCTCCGCCAAGTATGGTGAAGTGGTGGAAACGTGCCCCAGCGGTCAAACAATTCCGGG
TTCGATCGTCACGATCCAGATCAGCAACGGCATCCCGCCGGCTCCGCCTCCGCCACCGCTGCC
20 TGAGGATGGTGGGCCCGCCACCGCCGGTCGGATCGCAGGTGGTGGAGATTCCGGGGCTGCCGC
CGATCACCATTCCGCTGCTGGCGCCACCACCCCAAGCGCCTCCCCCGTAG

>Rv3721c dnaZX DNA polymerase III,[gamma] (dnaZ) and t (dnaX) TB.seq 4164995:4166728

MW:61892 >emb|AL123456|MTBH37RV:c4166728-4164992, dnaZX SEQ ID NO:138

25 GTGGCTCTTACCGCAAGTACCGACCGGCAAGCTTCGCGGAGGTGGTGGGGCAGGAGCACGT
CACCGCGCCGCTGTCGGTGGCGCTGGATGCCGGCCGGATCAACCACGCGTACCTGTTCTCTGG
GCCGCGTGGCTGCGGAAAGACGTGCTCAGCGCGTATCCTGGCGCGGTCGTTGAACTGTGCGCA
GGGCCCTACCGCAACCCGTCGCGGGTCTGCGAATCCTGCGTTTCGTTGGCGCCCAACGCCCC
CGGCAGCATCGACGTGGTAGAGCTGGATGCCGCCAGCCACGGCGGCGTGGACGACACCCGCG
30 AGCTGCGGGACCGCGCGTTCTATGCGCCGGTCCAGTCACGGTACCGGGTATTTATCGTCGACG
AGGCGCACATGGTGACCACCGCGGGATTCAACGCGCTGCTCAAGATCGTGGAGGAACCGCCC
GAACACCTGATCTTCATATTCGCCACCACCGAACCGGAGAAGGTAAGTACTGCCGACGATTCCGTCGC
GCACTCATCACTACCCGTTCCGGCTGCTGCCGCCGCGCACTATGCCGGCGGTTGCTCGCGCGGA
TCTGCGAGCAGGAGGGCGTCTGTCGACGATGCGGTGTACCCGTTGGTATCCGGGCCGCGC
35 GGAGGTTCCCCACGGGATACGCTCTCGGTGCTGGACCAATTGCTGGCTGGGGCCGCGGACAC
CCACGTGACCTACACCGGGCGCTGGGGCTGCTGGGTGTACCGACGTCGCCCTGATCGACG
ACGCGGTGACGCACTGGCCGCTTGCGATGCGGCCGCATTGTTCCGGGCGATCGAATCGGTGA

TCGATGGCGGACATGACCCTCGGCGTTTCGCTACCGATCTGCTGGAGCGATTCCGCGACCTGA
TTGTGCTGCAATCGGTTCCCGACGCGGCATCTCGCGGGTGGTGGATGCGCCCGAAGACGCG
CTGGATCGGATGCGCGAGCAAGCCGCCCGGATCGGGCGGGCGACCCTGACCCGATATGCCGA
GGTGGTGCAGGCCGGGCTAGGCGAGATGCGCGGTGCGACCGCGCCGCGTCTGCTGCTGGAA
5 GTGGTTTGCGCGCGACTGCTGCTGCCCTCGGCGAGCGACGCCGAATCGGCACTGTTGCAGCG
GGTCGAACGGATCGAGACCCGGTTGGACATGTCGATCCCGGCGCCGCAAGCCGTACCACGCC
CGTCGGCTGCGGCTGCCGAGCCGAAACACCAGCCCGCGCGTGAACCGAGACCGGTGCTGGCC
CCCACACCGGCCTCGAGCGAACCACCGTGGCCGCGGTTCTGGTCCATGTGGCCGACGGTGCG
CGACAAGGTGCGCCTGCGCAGCCGTACCACCGAGGTGATGCTGGCGGGTGCCACCGTCCGTG
10 CGCTAGAGGACAACACGCTGGTGTGACCCACGAATCGGCGCCGCTGGCGCGGCGGCTGTCC
GAACAGCGCAACGCCGATGTCCTCGCCGAGGCGCTTAAGACGCGCTGGGAGTCAACTGGCG
GGTGCGGTGTGAGACCGGTGAACCGGCTGCGGCGGCATCACCCGTCGGCGGGGGAGCGAAC
GTGGCGACCGCCAAGGCCGTAAACCCGCCCCACAGCGAATTCACCTCAGCGCGACGAAGAG
GAGCATGCTCGCCGAAGCCGGCCGTGGCGACCCGTCGCCGCGTCGCGACCCGGAAGAGGT
15 TCGACTCGAGCTGCTGCAGAACGAGCTGGGCGCGCGCCGGATAGACAACGCCTAG

>Rv3783 - TB.seq 4229255:4230094 MW:32337

>emb|AL123456|MTBH37RV:4229255-4230097, Rv3783 SEQ ID NO:139

ATGACATTCATGGATGCTCAAGCTAGCTTCCAGACACAGTCGCGGACACTGGCCCCGCGTCCGA
20 GGCGATCTGGTCGACGGGTTCCGCCGCCACGAGCTGTGGCTGCACCTGGGCTGGCAGGACAT
CAAGCAGCGGTACCGCCGCTCGGTGCTGGGGCCGTTCTGGATCACCATCGCCACCGGAACGA
CCGCCGTCGCGATGGGCGGCCTGTATTCCAAGCTGTTTCGGCTCGAGCTGTCTGAGCACCTGC
CCTACGTCACGCTCGGGCTGATCGTCTGGAACCTGATCAACGCCGCCATCCTGGACGGCGCAG
AGGTTTTCGTCGCCAACGAAGGTCTGATCAACAGCTGCCGGCACCGTTGAGCGTGCACGTCTA
25 TCGGTTGGTGTGGCGGCAGATGATCTTCTTCCCCACAACATCGTCATCTACTTCGTCATCGCG
ATCATCTTTCCTAAGCCGTGGTCGTGGGCGGATCTGTGTTTTCTTCCGGCGCTGGCGCTCATTT
TCCTCAATTGCGTTTGGGTGTCACTGTGTTTCGGCATCCTGGCGACCCGCTACCGCGACATCGG
CCCCGCTGCTGTTTTCCGTTGTGCAGTTGTTGTTCTTCATGACGCCGATCATCTGGAACGACGAGA
CCCTGCGTCGGCAGGGCGCGGGCCGCTGGTCGAGCATCGTCGAGCTCAACCCGCTGCTGCAC
30 TATCTGGACATCGTGCGGGCGCCACTGTTGGGCGCTCACCAGGAGCTGCGGCACTGGCTGGTG
GTGCTGGTGTGACCGTCGTGGCTGGATGCTGGCGGCGTTCCGATGCGGCAGTATCGCGC
GCGGGTGCCCTACTGGGTGTAG

>Rv3789 - TB.seq 4235371:4235733 MW:13378

35 >emb|AL123456|MTBH37RV:4235371-4235736, Rv3789 SEQ ID NO:140

ATGCGGTTCTGTTGTACCGGGCGGCCTCGCTGGGATAGTTGACTTTGGCCTCTACGTCGTGCTGT
ACAAGGTGGCGGGCCTACAGGTGACCTGTCCAAGGCCATCAGCTTCATCGTCGGCACCATCA

CCGCGTACCTGATCAACCGCCGGTGGACATTCCAGGCCGAGCCCAGCACGGCCCGATTCTGTCG
CGGTCATGCTCCTCTACGGAATCACCTTCGCCGTGCAGGTCGGAATCAACCACCTCTGCCTCGC
ACTCTTGCACTACCGGGCGTGGGCCATCCCCGTGCGTTTGTGATCGCGCAGGGCACCGCCAC
GGTAATCAACTTCATCGTGCAGCGAGCCGTGATCTTCCGGATCCGCTGA

5

>Rv3790 - TB.seq 4235776:4237158 MW:50164

>emb|AL123456|MTBH37RV:4235776-4237161, Rv3790 SEQ ID NO:141

ATGTTGAGCGTGGGAGCTACCACTACCGCCACCCGGCTGACCGGGTGGGGCCGCACAGCGCC
GTCGGTGGCGAATGTGCTTCGCACCCCAGATGCCGAGATGATCGTCAAGGCGGTGGCTCGGGT
10 CGCCGAGTCGGGGGGCGGCCGGGTGCTATCGCGCGCGGGCTGGGCCGCTCCTATGGGGAC
AACGCCCAAACGGCGGTGGGTTGGTGATCGACATGACGCCGCTGAACACTATCCACTCCATTG
ACGCCGACACCAAGCTGGTCGACATCGACGCCGGGTCAACCTCGACCAACTGATGAAAGCCG
CCCTGCCGTTCTGGGCTGTGGGTCCCGGTGCTGCCGGGAACCCGGCAGGTCACCGTCGGCGGG
GCGATCGCCTGCGATATCCACGGCAAGAACCATCACAGCGCTGGCAGCTTCGGTAACCACGTG
15 CGCAGCATGGACCTGCTGACCGCCGACGGCGAGATCCGTCATCTCACTCCGACCGGCGAGGA
CGCCGAAGTGTCTGGGCCACCGTCGGGGGCAACGGTCTCACCGGCATCATGCGGGCCAC
CATCGAGATGACGCCCACTTCGACGGCGTACTTCATCGCCGACGGCGACGTCACCGCCAGCCT
CGACGAGACCATCGCCCTGCACAGCGACGGCAGCGAAGCGCGCTACACCTATTCCAGTGCCTG
GTTGACGCGATCAGCGCTCCCCGAAGCTGGGCCGCGCGCGGTATCGCGTGGCCGCCTGG
20 CCACCGTCGAGCAATTGCCTGCGAACTGCGGAGCGAACCTTTGAAATTCGATGCGCCACAGCT
ACTTACGTTGCCCGACGTGTTTCCCAACGGGCTGGCCAACAAATATACCTTCGGCCCCGATCGGC
GAACTGTGGTACCGCAAATCCGGCACCTATCGCGGCAAGGTCCAGAACCTCACGCAGTTCTACC
ATCCGCTGGACATGTTCTGGCGAATGGAACCGCGCCTACGGCCCAGCGGGCTTCCTGCAATATC
AGTTCGTGATCCCCACAGAGGCGGTTGATGAGTTCAAGAAGATCATCGGCGTTATTCAAGCCTC
25 GGGTCACTACTCGTTTCTCAACGTGTTCAAGCTGTTGCGCCCCCGCAACCAGGCGCCGCTCAGC
TTCCCCATCCCGGGCTGGAACATCTGCGTCGACTTCCCCATCAAGGACGGGCTGGGGAAAGTTC
GTCAGCGAACTCGACCGCCGGGTACTGGAATTGCGCGGCCGGCTCTACACCGCCAAAGACTCC
CGTACCACCGCCGAAACCTTTCATGCCATGTATCCGCGCGTCGACGAATGGATCTCCGTGCGCC
GCAAGGTCGATCCGCTGCGCGTATTCGCCTCCGACATGGCCCGACGCTTGAGCTGCTGTAG

30

>Rv3791 - TB.seq 4237162:4237923 MW:27470

>emb|AL123456|MTBH37RV:4237162-4237926, Rv3791 SEQ ID NO:142

ATGGTCTTGATGCCGTAGGAAACCCCCAGACGGTGCTGCTGCTCGGTGGCACCTCCGAGATC
GGGCTCGCCATCTGCGAGCGCTACCTGCACAATTGCGCGGCCCGCATCGTGCTGGCCTGCCTG
35 CCCGACGACCCACGGCGGGAGGACGCGGCCGCTGCGATGAAGCAGGCCGGCGCGCGGTCCG
TGGAGCTGATCGACTTTGACGCCCTGGATACCGACAGCCACCCGAAGATGATCGAGGCGGCCT
TCTCCGGCGGTGATGTGGACGTGGCTATCGTCGCGTTCCGGCTTGCTCGGCGACGCCGAAGAGC

TGTGGCAGAACCAAGCGCAAGGCGGTGCAGATCGCCGAAATCAACTACACCGCAGCGGTTTCGG
TGGGCGTGCTGCTGGCTGAGAAGATGCGCGCTCAGGGCTTCGGTCAGATCATCGCGATGAGCT
CGGCCCGCCGGTGAGCGGGTGCGACGGGCGAACTTCGTCTACGGCTCCACCAAGGCCGGTCTG
GACGGGTTTTACCTGGGGTTGTGAGAAGCGCTGCGCGAGTACGGTGTTCTGTGTGCTGGTGATC
5 CGGCCCGGCCAGGTGCGTACCCGGATGAGCGCGCACCTCAAGGAAGCTCCATTGACCGTCGA
CAAGGAGTACGTGCGCAACCTCGCGGTGACCGCGTCCGCAAAAGGTAAGGAATTGTTTGGGC
GCCAGCAGCGTTCCGCTACGTCATGATGGTGTTGCGTCACATCCC CGGAGCATCTTCCGCAA
GCTGCCCATCTGA

10 >Rv3794 embA TB.seq 4243230:4246511 MW:115694
>emb|AL123456|MTBH37RV:4243230-4246514, embA SEQ ID NO:143
GTGCCCCACGACGGTAATGAGCGATCTACCGGATCGCACGCCTAGCAGCCGTCGTCTCGGGA
ATCGCGGGTCTGCTGCTGTGCGGCATCGTTCCGCTGCTTCCGGTGAACCAAACACCGCGACC
ATCTTCTGGCCGAGGGCAGCACCGCCGACGGCAACATCACCCAGATCACCGCCCCTCTGGTA
15 TCCGGGGCGCCACGCGCGCTGGACATCTCGATCCCCTGCTCGGCCATCGCCACGCTGCCCCG
CAACGGCGGCCTGGTGCTGTCCACACTGCCGGCCGGTGCGGTGGATACCGGTAAGGCCGGGC
TGTTCTGTCGCGCCAACCAAGACACGGTCTGCTGCGCTTCCGCGACTCGGTGGCCGCGGTG
GCGGCCCGCTCCACGATCGCAGCGGGAGGCTGTAGCGCGCTGCATATCTGGGCCGATACCGG
CGGCGCGGGCGCTGATTTTATGGGTATACCCGGCGGCGCCGGGACCCTGCCGCCGGAGAAGA
20 AGCCACAGGTTGGCGGCATCTTACCGACCTGAAGGTGCGAGCGCAGCCCCGGGCTGTGCGCC
CGCGTCGACATCGACACTCGGTTTATCAGACGCCCGGCGCGCTCAAGAAGGCCGTGATGCTC
CTCGGCGTGCTGGCGGTCCTGGTAGCCATGGTGCGGCTGGCCGCGCTGGACCGGCTCAGCAG
GGGCCGCAACCTGCGCGACTGGCTGACCCGATATCGCCCGCGGTGCGGGTGGATTGCCA
GCCGGCTCGCTGACGCGAGCGGTGATCGCGACCTTGTTGCTCTGGCATGTCATCGGCGCCACCT
25 CGTCCGATGACGGCTACCTTCTGACCGTCGCCCGGGTCGCCCGAAGGCCGGCTATGTAGCCA
ACTACTACCGGTATTTCCGGCACGACGGAGGCGCCGTTGCGACTGGTATACATCGGTGCTTGCCCA
GCTGGCGGCGGTGAGCACCGCGCGCTCTGGATGCGCCTGCCCGCCACCCTGGCCGGAATCG
CCTGCTGGCTGATCGTCAGCCGTTTCGTGCTGCGGCGGCTGGGACCGGGCCCGGGCGGGCTG
GCGTCCAACCGGGTCGCTGTGTTACCGCTGGTGCGGTGTTCTGTCCGCCTGGCTGCCGTTT
30 AACACGGCCTGCGTCCCGAGCCGCTGATCGCGCTGGGTGTGCTGGTCACGTGGGTGTTGGTG
GAACGGTCGATCGCGCTCGGACGGCTGGCCCCGGCGCGGTAGCCATCATCGTGCGGACGCT
TACCGCGACGCTGGCACCGCAGGGGTTGATCGCGCTGGCCCCGCTGCTGACTGGTGCGCGCG
CCATCGCCCAGAGGATCCGGCGCCCGCGGGCGACCGATGGACTGCTGGCGCCGCTGGCGGT
GCTGGCCGCGGCGTTGTGCGCTGATACCGTGGTGGTGTTCGGGACCAGACGCTGGCCACGGT
35 GGCCGAATCGGCACGCATCAAGTACAAGGTGCGCCCGACCATCGCCTGGTACCAGGACTTCCT
GCGCTACTACTTCCTTACCGTGGAGAGCAACGTTGAGGGTTCGATGTCCCGCCGTTTCGCGT
GCTGGTGTGCTGTTCTGCCTGTTCCGGGTGCTGTTCTGTGCTGCTGCGCGCGGCCGGGTGGC

GGGGCTGGCCAGCGGCCCGGCCTGGCGACTGATCGGCACTACGGCGGTTCGGCCTGCTGCTGC
TCACGTTACGCCAACCAAGTGGGCCGTGCAGTTCGGCGCATTGCCGGGCTGGCCGGGGTGT
TGGGTGCGGTACCGCGTTCACCTTTGCCCGCATCGGTCTACATAGTCGACGCAACCTCACGCT
GTACGTGACCGCGTTGCTGTTCTGTGCTGGCGTGGGCAACCTCGGGCATCAACGGGTGGTTCTA
5 CGTCGGCAACTACGGGGTGCCGTGGTATGACATCCAGCCCGTCATCGCCAGCCACCCGGTGAC
GTCGATGTTTCTGACGCTGTGATCCTCACCGGATTGCTGGCAGCCTGGTATCACTTCCGGATG
GACTACGCCGGGCACACCGAAGTCAAAGACAACCGGCGCAACCGCATCTTGGCCTCTACGCCA
CTGCTGGTGGTCGCGGTGATCATGGTCGCAGGCGAAGTCGGCTCGATGGCCAAGGCCGCGGT
GTTCCGTTACCCGCTTTACACCACCGCCAAGGCCAACCTGACCGCGCTCAGCACCGGGCTGTC
10 CAGCTGTGCGATGGCCGACGACGTGCTGGCCGAGCCCGACCCCAATGCCGGCATGCTGCAAC
CGGTTCCGGGCCAGGCGTTTCGGACCGGACGGACCGCTGGGCGGTATCAGTCCCGTCGGCTTC
AAACCCGAGGGCGTGGGCGAGGACCTCAAGTCCGACCCGGTGGTCTCCAAACCCGGGCTGGT
CAACTCCGATGCGTCGCCAACAAACCCAACGCCGCCATCACCGACTCCGCGGGCACCGCCGG
AGGGAAGGGCCCGGTTCGGGATCAACGGGTTCGCACGCGGCGCTGCCGTTTCGGATTGGACCCGG
15 CACGTACCCCGGTGATGGGCAGCTACGGGGAGAACAACTGGCCGCCACGGCCACCTCGGCC
TGGTACCAGTTACCGCCCCGACGCCCGACCGGCCGCTGGTGGTGGTTTCCGCGGCCGGCGC
CATCTGSTCCTACAAGGAGGACGGCGATTTCATCTACGGCCAGTCCCTGAAACTGCAGTGGGG
CGTCACCGGGCCCGGACGGCCGCATCCAGCCACTGGGGCAGGTATTTCCGATCGACATCGGACC
GCAACCCGCGTGCGCAATCTGCGGTTTCCGCTGGCCTGGGCGCCGCCGGAGGCCGACGTGG
20 CGCGCATTGTGCCTATGACCCGAACCTGAGCCCTGAGCAATGGTTCGCCTTACCCCCGCCCC
GGGTTCCGGTGCTGGAATCTCTGCAGCGTTGATCGGGTCAGCGACACCGGTGTTGATGGACA
TCGCGACCGCAGCCAACCTCCCTGCCAGCGACCGTTTTCCGAGCATCTCGGCATTGCCGAGC
TTCCGCAGTACCGGATCCTGCCGGACCACAAGCAGACGGCGGCGTCGTCGAACCTATGGCAGT
CCAGCTCGACCGGCGGTCCGTTCTGTTACCCAGGCGCTGCTGCGCACCTCGACGATCGCCA
25 CGTACCTGCGTGGGGACTGGTATCGCGACTGGGGATCGGTGGAGCAGTACCACCGGCTGGTG
CCGGCCGATCAGGCTCCAGACGCCGTTGTCGAGGAGGGCGTGATCACTGTGCCCGGCTGGGG
TCGGCCAGGACCGATCAGGGCGCTGCCATGA

>Rv3795 embB TB.seq 4246511:4249804 MW:118023

30 >emb|AL123456|MTBH37RV:4246511-4249807, embB SEQ ID NO:144

ATGACACAGTGCGCGAGCAGACGCAAAAGCACCCCAATCGGGCGATTTTGGGGGCTTTTGC
TCTGCTCGCGGGACGCGCTGGGTGGCCACCATCGCCGGGCTGATTGGCTTTGTGTTGTCGGTG
GCGACGCCGCTGCTGCCCGTCGTGCAGACCACCGCGATGCTCGACTGGCCACAGCGGGGGCA
ACTGGGCAGCGTGACCGCCCCGCTGATCTCGCTGACGCCGGTCGACTTTACCGCCACCGTGCC
35 GTGCGACGTGGTGCGCGCCATGCCACCCGCGGGCGGGGTGGTGTGGGCACCGCACCCAAG
CAAGGCAAGGACGCCAATTTGCAGGCGTTGTTCTGTCGTGTCAGCGCCAGCGCGTGGACGTC
ACCGACCGCAACGTGGTGTCTTGTCCGTGCCGCGCAGCAGGTGACGTCCCCGAGTGTCAA

CGCATCGAGGTCACCTCTACCCACGCCGGCACCTTCGCCAACTTCGTGCGGGCTCAAGGACCCG
TCGGGCGCGCCGCTGCGCAGCGGCTTCCCCGACCCCAACCTGCGCCCGCAGATTGTCGGGGT
GTTACCCGACCTGACCGGGCCCGCGCCGCCGGGCTGGCGGTCTCGGCGACCATCGACACCC
GGTCTCCACCCGGCCGACCACGCTGAAACTGCTGGCGATCATCGGGGCGATCGTGGCCACCG
5 TCGTCGCACTGATCGCGTTGTGGCGCCTGGACCACTTGGACGGGCGGGGCTCAATTGCCAGC
TCCTCCTCAGGCCGTTCCGGCCTGCATCGTCGCCGGCGGCATGCGCCGGCTGATTCCGGCAA
GCTGGCGCACCTTCACCTGACCGACGCCGTGGTGATATTCGGCTTCCTGCTCTGGCATGTCAT
CGGCGCGAATTTCGTGCGACGACGGCTACATCCTGGGCATGGCCCGAGTCGCCGACCACGCCG
GCTACATGTCCAATAATTTCCGCTGGTTCGGCAGCCCGGAGGATCCCTTCGGCTGGTATTACAA
10 CCTGCTGGCGCTGATGACCCATGTCAGCGACGCCAGTCTGTGGATGCGCCTGCCAGACCTGGC
CGCCGGGCTAGTGTGCTGGCTGCTGCTGTCGCGTGAGGTGCTGCCCCGCCTCGGGCCGGCGG
TGGAGGCCAGCAAACCCGCCTACTGGGCGGCGGCCATGGTCTTGCTGACCGCGTGGATGCCG
TTCAACAACGGCCTGCGGCCGGAGGGCATCATCGCGCTCGGCTCGCTGGTCACCTATGTGCTG
ATCGAGCGGTCCATGCGGTACAGCCGGCTCACACCGGCGGCGCTGGCCGTGTTACCGCCGC
15 ATTCACACTGGGTGTGCAGCCACCGGCCTGATCGCGGTGGCCGCGCTGGTGGCCGGCGGCC
GCCCGATGCTGCGGATCTTGGTGCGCCGTATCGCCTGGTCGGCACGTTGCCGTTGGTGTGCG
CGATGCTGGCCGCCGGCACCGTATCCTGACCGTGGTGTTCGCCGACCAGACCCTGTCAACGG
TGTTGGAAGCCACCAGGGTTCGCGCCAAAATCGGGCCGAGCCAGGCGTGGTATACCGAGAACC
TGCGTTACTACTACCTCATCCTGCCACCGTCGACGGTTCGCTGTGCGGCGCTTCGGCTTTTT
20 GATCACCGCGCTATGCCTGTTACCGCGGTGTTTCATCATGTTGCGGCGCAAGCGAATTCACAGC
GTGGCCCGCGGACCGGCGTGGCGGCTGATGGGCGTCATCTTCGGCACCATGTTCTTCCTGATG
TTCACGCCACCAAGTGGGTGCACCACTTCGGGCTGTTCCGCCCGGTAGGGGCGGCGATGGC
CGCGCTGACGACGGTGTTGGTATCCCATCGGTGCTGCGCTGGTCGCGCAACCGGATGGCGTT
CCTGGCGGCGTTATTTCTCCTGCTGGCGTTGTGTTGGGCCACCACCAACGGCTGGTGGTATGTC
25 TCCAGCTACGGTGTGCCGTTCAACAGCGCGATGCCGAAGATCGACGGGATCACAGTCAGCACA
ATCTTTTTCGCCCTGTTTGCGATCGCCGCCGGCTATGCGGCCTGGCTGCACTTCGCGCCCCGC
GGCGCCGGCGAAGGGCGGCTGATCCGCGCGCTGACGACAGCCCCGGTACCGATCGTGGCCG
GTTTCATGGCGGCGGTGTTTCGTGCGCTCCATGGTGGCCGGGATCGTGCGACAGTACCCGACCT
ACTCCAACGGCTGGTCCAACGTGCGGGCGTTTGTGCGCGGCTGCGGACTGGCCGACGACGTA
30 CTCGTGAGCCTGATACCAATGCGGGTTTCATGAAGCCGCTGGACGGCGATTGCGGTTCTTGG
GGCCCCCTTGGGCCCGCTGGGTGGAGTCAACCCGGTCGGCTTCACGCCCAACGGCGTACCGGA
ACACACGGTGGCCGAGGCGATCGTGATGAAACCCAACGACCCGGCACCAGCTACGACTGGGA
TGCGCCGACCAAGCTGACGAGTCCTGGCATCAATGTTTCTACGGTGCCGCTGCCCTATGGGCT
CGATCCCGCCCGGGTACCGTTGGCAGGCACCTACACCACCGGCGCACAGCAACAGAGCACACT
35 CGTCTCGGCGTGGTATCTCCTGCCTAAGCCGGACGACGGGCATCCGCTGGTCGTGGTGACCGC
CGCGGGCAAGATCGCCGGCAACAGCGTGCTGCACGGGTACACCCCGGGCAGACTGTGGTG
TCGAATACGCCATGCCGGGACCCGGAGCGCTGGTACCGCGCGGGCGGATGGTGCCCGACGAC

CTATACGGAGAGCAGCCCAAGGCGTGGCGCAACCTGCGCTTCGCCCAGCAAAGATGCCCGC
CGATGCCGTCGCGGTCCGGGTGGTGGCCGAGGATCTGTGCTGACACCGGAGGACTGGATCG
CGGTGACCCCGCCGCGGTACCGGACCTGCGCTCACTGCAGGAATATGTGGGCTCGACGCAG
CCGGTGCTGCTGGA CTGGGCGGTTCGGTTTGGCCTTCCCGTGCCAGCAGCCGATGCTGCACGC
5 CAATGGCATCGCCGAAATCCCGAAGTTCGCGATCACACCGGACTACTCGGCTAAGAAGCTGGAC
ACCGACACGTGGGAAGACGGCACTAACGGCGGCCTGCTCGGGATCACCGACCTGTTGCTGCG
GGCCACGTCATGGCCACCTACCTGTCCCGCGACTGGGCCCCGCGATTGGGGTTCCCTGCGCAA
GTTCGACACCCTGGTCGATGCCCTCCCGCCAGCTCGAGTTGGGCACCGCGACCCGCAGCG
GCCTGTGGTCACCGGGCAAGATCCGAATTGGTCCATAG

10

>Rv3834c serS seryl-tRNA synthase TB.seq 4307655:4308911 MW:45293

>emb|AL123456|MTBH37RV:c4308911-4307652, serS SEQ ID NO:145

GTGATCGACCTGAAGCTGCTTCGTGAAAACCCGACGCGGTACGCCGCTCACAACCTCAGCCGC
GGCGAGGACCCGGCGCTGGTAGATGCCCTGCTGACGGCCGACGCCGCCCGCCGGGCCGTGA
15 TCTCGACCGCCGATTCTGTTACGGGCCGAGCAGAAAGCCGCCAGCAAAGCGTGGGTGGCGCG
TCTCCCGAAGAGCGCCCGCCGCTGCTGCGGCGCGCGAAGGAACTCGCCGAGCAGGTCAAAGC
CGCTGAGGCCGACGAGGTGGAAGCGGAGGCGGCGTTACCGCGGCGCACCTGGCGATCTCGA
ATGTCATCGTGGACGGGGTACCCGCCGGCGGGGAGGACGACTACGCGGTGCTCGACGTCGTC
GGCGAGCCCAGCTACCTCGAGAACCCCAAGGACCACCTGGAGCTCGGCGAGTCGCTGGGCCT
20 GATCGACATGCAGCGCGGCCGAAGGTGTCGGGTTCACGGTTCTACTTCCTGACCGGTTCGGGG
TGCCCTACTGCAGCTTGATTGCTGCAGCTGGCGCTGAAGCTAGCCGTCGACAACGGCTTTGTC
CCTACGATCCCGCCGGTGTGCTGGTGCGCCCGGAAGTGATGGTAGGCACGGGATTTCTAGGCGCC
CACGCCGAGGAGGTGTACCGGGTAGAGGGCGACGGCCTCTACCTTGTGGGCACCTCCGAGGT
ACCGCTGGCGGGGTATCACTCCGGCGAGATTCTGGACCTTTCGCGGGGCCGCTGCGGTATGC
25 GGGCTGGTCGTCGTGTTTCCGACGTGAGGCCGGCAGCCATGGCAAGGACACGCGCGGCATCA
TCCGGGTGCACCAAGTTCGACAAAGTCGAGGGCTTCGTCTACTGCACACCGGCCGACGCGGAGC
ACGAACATGAGCGGCTGCTGGGCTGGCAGCGCCAGATGCTGGCACGCATCGAGGTGCCGTAT
CGGGTCATCGACGTGGCCGCGGGTGATCTCGGCTCGTCGGCCGCCCGCAAGTTCGACTGCGA
GGCGTGGATTCCGACGCAGGGGGCCTATCGCGAGCTGACGTGACGTGCAACTGCACCACCTT
30 TCAGGCGCGCCGTTGGCGACCCGCTACCGGGATGCCAGCGGCAAGCCGCAGATCGCGGCCA
CCCTCAACGGAACGCTGGCCACCACCCGGTGGCTGGTTGCGATCCTGGAGAACCACCAGCGG
CCCGACGGCAGCGTTAGAGTCCCGGACGCACTGGTTCCGTTCGTGGGTGTCGAAGTGCTGGAG
CCGGTCGCTTAG

35

>Rv3907c pcnA polynucleotide polymerase TB.seq 4391631:4393070 MW:53057

>emb|AL123456|MTBH37RV:c4393070-4391628, pcnA SEQ ID NO:146

GTGCCGGAAGCCGTCCAGGAAGCCGATCTGCTAACCGCCGCTGCGGTTGCCTTGAACAGGCAT
 GCTGCCTTATTGCGGGAACTCGGGTCGGTGTTCCGCCGCCGCGGGACACGAGTTGTATCTGGTC
 GGCGGTTCCGGTGCGAGATGCACTGTTGGGCCGGTTGAGCCCCGACCTGGACTTCACCACCGAC
 GCCCCGTCCCGAGCGGGTGCCAGGAGATCGTGCGGCCGTGGGCCGATGCGGTGTGGGATACCG
 5 GAATCGAATTCGGCACCGTCGGCGTGGGTAAGAGCGACCACCGCATGGAGATCACCACATTCC
 GTGCCGACAGCTACGACCGGGTTTCGCGTCATCCAGAGGTACGTTTCGGCGATTGCCTCGAGG
 GCGATCTGGTCCGCCGCGACTTCACCACGAACGCAATGGCTGTGCGCGTCACCGCCACTGGGC
 CGGGCGAATTCCTGGATCCGCTTGGTGGCTTGGCGGCGCTGCGGGCCAAGGTGTTAGACACCC
 CGGCGGCGCCGTGCGGGTCCTTTGGCGACGATCCGTTGCGGATGCTGCGCGCCGCGCGGTTTC
 10 GTCTCGCAACTTGATTGCGGGTGGCGCCGCGGGTCCGCGCGGCGATCGAAGAGATGGCGCC
 GCAGTTGGCCCGAATCAGCGCCGAACGGGTGGCCGCCGAGCTGGACAAGCTGCTGGTGGTG
 AGGATCCGGCCGCGGGTATCGACCTGATGGTGCAGAGCGGTATGGGTGCTGTGGTCTTGCCTG
 AAATCGGTGGGATGCGGATGGCGATCGACGAACATCACCAGCACAAGGACGTCTATCAGCATT
 CTTGACCGTGCTGCGGCAGGCGATCGCGCTGGAGGACGACGGCCCGGATCTGGTGTTCGCT
 15 GGGCGGCGCTGCTGCACGACATCGGCAAGCCCGCCACCCGCGTCACGAACCCGACGGTGGG
 GTGAGCTTCCATCACCACGAAGTGGTGGCGCCAAGATGGTGGCAAGCGGATGCGGGGCGCT
 GAAGTATTCGAAGCAGATGATCGACGACATCTCGCAGCTGGTCTACCTGCATCTGCGGTTTCAC
 GGCTACGGCGATGGGAAATGGACCGACTCTGCGGTGCGCCGCTATGTCACCGACGCGGGGGC
 CCTACTGCCACGGCTGCACAAGCTGGTGGCGCGCGACTGCACGACCCGCAACAAGCGCCGGG
 20 CCGCGCGGTTGCAGGCCAGTTACGACCGGCTGGAAGAGCGGATCGCGGAGCTGGCCGCCAG
 GAGGATCTGGATCGGGTGGCCCCGACCTGGACGGCAACCAGATCATGGCGGTGCTCGACATT
 CCGGCGGGGCCCCGAAGTCGGCGAGGCGTGGCGCTACTTGAAGGAGCTGCGGCTAGAGCGCG
 GCCCCGTTGTCCACCGAGGAGGCGACAACCGAGCTGCTGTCCTGGTGGAATCACGGGGGAAC
 CGCTAG

25

TABLE 4

>Rv0002 dnaN DNA polymerase III, b-subunit TB.seq 2052:3257 MW:42114 SEQ ID NO:147

MDAATTRVGLTDLTFRLLRESFADAVSWVAKNLPARPAVPVLSGVLLTGSDNGLTISGFDYEVSAEA
 QVGAEIVSPGSVLVSGRLLSDITRALPNKPVDVHVEGNRVALTCGNARFSLPTMPVEDYPTLPTLPEE
 30 TGLLPAELFAEAIQVAIAAGRDDTLPLMTGIRVEILGETVVLAATDRFRLAVRELKWSASSPDIEAAVL
 VPAKTLAEAAKAGIGGSDVRLSLGTGPGVGKDGLLGISGNGKRSTTRLLDAEFPKFRQLLPTEHTAVA
 TMDVAELIEAIKLVALVADRGAVRMEFADGSVRLSAGADDVGRAEEDLVVDYAGEPLTIAFNPTYLT
 DGLSSLRSERVSFGFTTAGKPALLRPVSGDDRPVAGLNGNGPFPVAVSTDYVYLLMPVRLPG

35 >Rv0003 recF DNA replication and SOS induction TB.seq 3280:4434 MW:42181 SEQ ID NO:148
 VYVRHLGLRDFRSWACVDLELHPGRTVFVGPNGYGKTNLIEALWYSTTLGSHRVSADLPLIRVGTDR
 AVISTIVNDGRECAVDLEIATGRVKNKARLNRSSVRSTRDVVGVLRVLFAPEDLGLVRGDPADRRR

YLDDLAIVRRPAIAAVRAEYERVLQRQTALLKSVPGARYRGDRGVFDTLEVWDSRLAEHGAELVAARI
DLVNQLAPEVKKAYQLLAPESRSASIGYRASMDVTGPSEQSDIDRQLLAARLLAALAARRDAELERG
VCLVGPHRDDLILRLGDQPAKGFAWSLAVALRLAAYQLLRVDGGEPVLLDDVFAELDMRR
RALATAAESAEQVLVTAAVLEDIPAGWDARRVHIDVRADDTGSMSVLP

5

>Rv0005 gyrB DNA gyrase subunit B TB.seq 5123:7264 MW:78441 SEQ ID NO:149

MGKNEARRSALAPDHGTVCPLRRLNRMHATPEESIRIVAAQKKKAQDEYGAASITILEGLEAVRKR
PGMYIGSTGERGLHHLIWEVVDNAVDEAMAGYATTNNVWLEDGGVEVADDGRGIPVATHASGIPTV
DVVMTQLHAGGKFDSDAYAISGGLHGVGVSVVNALSTRLEVEIKRDGYEWSQVYEKSEPLGLKQGA
10 PTKKTGSTVRFWADPAVFETTEYDFETVARRLQEMAFLNKGLTINLTDERTQDEVVDEVSDVAEA
PKSASERAAESTAPHKVKSRFTFHYPGGLVDFVKHINRTKNAIHSSIVDFSGKGTGHEVEIAMQWNAG
YSESVHTFANTINHEGGTHEEGFRSALTSVNNKYAKDRKLLKDKDPNLTGDDIREGLAAVISVKVSE
PQFEGQTKTKLGNTVEKSFVQKVCNEQLTHWFEANPTDAKVVVNKAVSSAQARIAARKARELVRRK
SATDIGGLPGKLADCRSTDPRKSELYVVEGDSAGGSAGKSGRDSMFQAILPLRGKIINVEKARIDRVLK
15 NTEVQAIITALGTGIHDEFDIGKLRYHKIVLMADADVGGHISTLLLTLFRFMRPLIENGHVFLAQPPLY
KLKWQRSDFEFAYSRRERDGLLEAGLKAGKKINKEDGIQRYKGLGEMDAKELWETTMPSVRVLRQ
VTLDAAAADLFSILMGEDVDARRSFITRNAKDVRFLDV

>Rv0006 gyrA DNA gyrase subunit A TB.seq 7302:9815 MW:92276 SEQ ID NO:150

MTDTTLPDDSLDRIEVDIEQEMQRSYIDYAMSVIVGRALPEVRDGLKPVHRRVLYAMFDSGFRPD
RSHAKSARSVAETMGNYHPHGDAISIYDSLVRMAQPWSLRYPLVDGQGNFGSPGNDPPAAMRYTEA
RLTPLAMEMLREIDEETVDFIPNYDGRVQEPTVLPSPFPNLLANGSGGIAGMATNIPPNNLRELADA
VFWALENHDADEEETLAAVMGRVKGPDPFTAGLIVGSQGTADAYKTGRGSIRMRGVVEVEEDSRG
RTSLVITELPYQVNHDFITISIAEQVRDGLAGISNIEDQSSDRVGLRIVIEIKRDAVAKVWINNLYKHTQ
25 LQTSFGANMLAIVDGVPRTRLRLDQLIRYVDHQLDVIVRRTTYRLRKANERAHILRGLVKALDALDEVI
ALIRASETVDIARAGLIELLDIDEIQAQAILDMQLRRLAALERQRIIDDLAKIEAEIADLEDILAKPERQGI
VRDELAIEIVDRHGDDRRTRIIAADGDVSDIEDLIAREDVVVTITETGYAKRTKTDLYRSQKRGKGKVQG
AGLKQDDIVAHFFVCSTHDLILFFTQGRVYRAKAYDLPEASRTARGQHVANLLAFQPEERIAQVIQIR
GYTDAPYLVLATRNLVKKSKLTDSDNRSGGIVAVNLRDNDELVGAVLCSAGDLLLLVSANGQSIR
30 FSATDEALRPMGRATSGVQGMRFNIDRLLSLNVVREGTYLLVATSGGYAKRTAIEEYPVQGRGKG
GVLTVMYDRRRGRLVGALIVDDDSELYAVTSGGGVIRTAARQVRKAGRQTKGVRLMNLGEGDTLLAI
ARNAEESGDDNAVDANGADQTGN

>Rv0014c pknB serine-threonine protein kinase TB.seq 15593:17470 MW:66511 SEQ ID NO:151

MTTPSHLSDRYELGEILGFGGMSEVHLARDLRLHRDVAVKVLRADLARDPSFYLRFRREAQNAALN
HPAIVAVYDTGEAETPAGPLPYIVMEYVDGVTLRDIVHTEGPMTPKRAIEVIADACQALNFSHQNGIIH
RDVKPANIMISATNAVKVMDFGIARIAADSGNSVTQTAAVIGTAQYLSPEQARGDSVDARSDVYSLGC

VLYEVLTEGPPFTGDSPPSVAYQHVEDPIPPSARHEGLSADLDAVVLKALAKNPENRYQTAAEMRA
 DLVRVHNGEPPEAPKVLTAERTSLLSSAAGNLGSPRTDPLPRQDLDDTDORRSIGSVGRWVAVVA
 VLAULTVWVTIAINTFGGITRDVQVPDVRGQSSADAIATLQNRGFKIRTLQKPDSTIPPDHVIGTDPAAN
 TSVSAGDEITVNVSTGPEQREIPDVSTLYAEAVKKLTAAGFGRFKQANSPTPELVGKVIGTNPPAN
 5 QTSAITNVIIIIVGSGPATKDIPDVAGQTVDAQKNLNVYGFTKFSQASVDSPPRAGEVTGTNPPAGT
 TVPVDSVIELQVSKGNQFVMPDLSGMFWVDAEPRLRALGWTGMLDKGADV DAGGSQHNRVYQN
 PPAGTGVNRDGIITLRFQG

>Rv0016c pbpA TB.seq 18762:20234 MW:51577 SEQ ID NO:152

10 MNASLRRISVTVMALIVLLLLNATMTQVFTADGLRADPRNQRVLLDEYSRQRGQITAGGQLLAYSAT
 DGRFRFLRVYPNPEVYAPVTGFYSLRYSSTALERAEDPILNGSDRRLFGRRADFFTGDRDPRGGNV
 DTTINPRIQQAGWDAMQQGCGYGPCKGAVVALEPSTGKILALVSSPSYDPNLLASHNPEVQAQAWQR
 LGDNPASPLTNRAISETYPPGSTFKVITTAALAAAGATETEQLTAAPTIPLPGSTAQLENYGGAPCGDE
 PTVSLREAFVKSCNTAFVQLGIRTGADALRSMARAFGLDSPRPPTPLQVAESTVGPIPDAAALGMTSI
 15 GQKDVALTPLANAIEAATIANGGITMRPYLVGSLKGPDLANISTTVGYQQRRRAVSPQVAAKLTLMVG
 AEKVAQQKGAIPGVQIASKTGTAEHGTDPRHTPPHAWYIAFAPAQAPKVAVAVLVENGADRLSATGG
 ALAAPIGRAVIEAALQGEP

>Rv0017c rodA TB.seq 20234:21640 MW:50612 SEQ ID NO:153

20 MTTRLQAPVAVTPPLPTRRNAELLLLCAAVITFAALLVQANQDQGVWDLTSGLAFLTLFGSAHL
 AIRRFAPYTDPLLLPVVALLNGLGLVMHRLDLVDNEIGEHRHPSANQQMLWTLVGVAAFALVVTFLK
 DHRQLARYGYICGLAGLVFLAVPALLPAALSEQNGAKIWIRLPGFSIQPAEFSKILLIFFSAVLVAKRG
 LFTSAGKHLLGMTLPRPRDLAPLLAAWVISVGMVFEEKDLGASLLLYTSFLVWVYLATQRFSSWVIGL
 TLFAAGTLVAYFIFEHVRRLRVQTWLDPFADPDGTGYQIVQSLFSFATGGIFGTGLNGGQPDTPAAST
 25 DFIIAAGEELGLVGLTAILMLYTIIRGLRTAIATRDSFGKLLAAGLSSTLAIQLFIWGGVTRLIPLTGLT
 TPWMSYGGSSLLANYILLAILARISHGARRPLRTRPRNKSPITAAGTEVIERV

>Rv0018c ppp TB.seq 21640:23181 MW:53781 SEQ ID NO:154

30 VARVTLVRLYAARS DRGLVRANNEDSVYAGARLLALADGMGGHAAGEVASQLVIAALAHLDDEPG
 GDLLAKLDAAVRAGNSAIAAQVEMEPDLEGMGTTLTAILFAGNRLGLVHIGDSRGYLLRDGELTQITK
 DDTFVQTLVDEGRITPEEAHSHPQRSIMRALTGHEVEPTLTMREARAGDRYLLCSDGLSDPVSDETI
 LEALQIPEVAESAHRILALRGGGPDNVTWVADVVDYDYGQTQPILAGAVSGDDDQLTLPNTAAG
 RASAIQRKEIVKRVPQADTFSRPRWSGRRLAFVVALVTLMTAGLLIGRAIRSNYYVADYAGSVSI
 MRGIQGSLLGMSLHQPYLMGCLSPRNELSQISYGQSGGPLDCHLMKLEDLRPPERAQVRAGLPAGT
 35 LDDAIGQLRELAANSLPPCPAPRATSPPGRPAPPTTSETTEPNVTSSPASPSPTTSAPAPTGTTPAIP
 TSASPAAPASPTTPWPVTSSPTMAALPPPPQPGIDCRAAA

>Rv0019c - TB.seq 23273:23737 MW:17153 SEQ ID NO:155

MQQLVLQLTRAGFLMLLWFIWSVLRLKTDIYAPTGAVMMRRGLALRGTLGARQRRHAARYLVVT
EGALTGARITLSEQPVLIGRADDSTLVLTDYASTRHARLSMRGSEWYVEDLGSTNGTYLDRAKVTT
AVRVPIGTPVRIGKTAIELRP

5

>Rv0020c - TB.seq 23864:25444 MW:56881 SEQ ID NO:156

MGSQKRLVQRVERKLEQTVGDAFARIFGGSIVPQEEVALLRREAADGIQSLQGNRLLAPNEYIITLGV
HDFEKLGADELKSTGFARDLADYIQEQGWQTYGDVVRFEQSSNLHTGQFRARGTVNPDVETHP
PVIDCARPQSNHAFGAEPGVAPMSDNSSYRGQGQGRPDDEYDDRYARPQEDPRGGPDPQGGG
10 DPRGGYPPEGGYPPQPGYPRRHPDQGDYPEQIGYPDQGGYPEQRGYPEQRGYPDQRGYQDQ
GRGYPDQGGQGGYPPPYEQRPVSPGPAAGYGAPGYDQGYRQSGGYGSPGGGQPGYGGYGEY
GRGPARHEEGSYVPSGPPGPPEQRPAYPDQGGYDQGYQQGATTYGRQDYGGGADYTRYTESPR
VPGYAPQGGGYAEPAGRDYDYGQSGAPDYGQPAPGGYSGYGQGGYGSAGTSVTLQLDDGSGRT
YQLREGSNIIGRGQDAQFRLPDTGVSRRHLEIRWDGQVALLADLNSTNGTTVNNAPVQEWQLADGD
15 VIRLGHSEIIVRMH

>Rv0032 bioF2 C-terminal similar to B. subtilis BioF TB.seq 34295:36607 MW:86245

SEQ ID NO:157

MPTGLGYDFLRPVEDSGINDLKHYFMA DLADGQPLGRANLYSVCFDLATTDRKLT PAWRTTIKRW
20 PGFMTFRFLECGLLTMVSNPLALRS DTDLERVLPLAGQMDQLAHDDGSDFLMIRDVDPEHYQRYL
DILRPLGFRPALGFSRVDTTISWSSVEEALGCLSHKRRLPLKTSLEFRERFGIEVEELDEYAEHAPVLA
RLWRNVKTEAKDYQREDLNPEFFAACSRHLHGRSRLWLFYRQGTPIAFFLNWWGADENYILLEWGI
DRDFEYHRKANLYRAALMSLKDAISRDKRRMEMGITNYFTKL RIPGARVIPTIYFLRHSTDPVHTATL
ARMMMHNIQRPTLPDDMSEEF CRWEERIRLDQDGLPEHDIFRKIDRQH KYTGLKLGGVYGFYPRFT
25 GPQRSTVKAELGEIVLLGTNSYLGLATHPEVVEASAEATRRYGTGCSGSPLLNGTLDLHVSLEQEL
ACFLGKPAAVLCSTGYQSNLAAISALCESGDMIIQDALNHRSLFDAARLSGADFTLYRHNDMDHLARV
LRRTEGRRRIIVDAVFSMEGTVADLATIAELADRHGCRVYVDESHALGVLGPDGRGASAAALGVLAR
MDVVMGTFSKSFASVGGFIAGDRPVVDYIRHNGSGHVF SASLPPAAAAATHAALRVSRREPDRRAR
VLAAAEYMATGLARQGYQAEYHGTAIVPVILGNPTVAHAGYLRLMRSGVYVNPVAPPVPEERSGFR
30 TSYLADHRQSDLDRALHVFAGLAEDLTPQGAAL

>Rv0050 ponA1 TB.seq 53661:55694 MW:71119 SEQ ID NO:158

VVILLPMVFTTMAYLIVDPKPGDIRTNQVSTILASDGSEIAKIVPEGNRVDVNL SQVPMHVRQAVIAA
EDRNFYSNPGFSFTGFARAVKNNLFGGDLQGGSTITQQYVKNALVGS AQHGW SGLMRKAKELVIAT
35 KMSGEW SKDDVLQAYLNIIYFGRGAYGISAASKAYFDKPVEQLTVAEGALLAALIRRPSTLDPVADPE
GAHARWNWVLDGMVETKALSPNDRAAQVFPETVPPDLARAENQTKGPNGLIERQVTRELLELFNID
EQLNTQGLVTTTIDPQAQRAAEKAVAKYLDGQDPDMRAAVVSIDPHNGAVRAYYG GDNANGFDF

AQAGLQTGSSFKVFALVAALEQQIGLGYQVDSSPLTVDGKITNVEGEGCGTCNIAEALKMSLNTSY
RLMLKLNGGPQAVADAAHQAGIASSFPQVAHTLSEDGKGGPPNNGIVLGQYQTRVIDMASAYATLAA
SGIYHPPHFVQKVVSANGQVLFDASTADNTGDQRIPKAVADNVTAAMEPIAGYSRGHNLAGGRDSA
AKTGTTQFGDTTANKDAWMVGYTPSLSTAVWVGTVKGDEPLVTASGAAIYGSGLPDSIWKATMDGA
5 LKGTNETFPKPTTEVGGYAGVPPPPPPPEVPPSETVIQPTVEIAPGITIPGPPTTITLAPPPPPAPPAAT
PTPPP

>Rv0051 - TB.seq 55694:57373 MW:61210 SEQ ID NO:159
VTGALSQSSNISPLPLAADLRSADNRDCPSRTDVLGAALANVGGPVGRHALIGRTRLMTPLRVMFAI
10 ALVFLALGWSTKAACLQSTGTGPGDQRVANWDNQRAYYQLCYSDTVPLYGAELLSQGKFPYKSSWI
ETDSNGTPQLRYDGGQIAVRYMEYPVLTGIYQYLSMAIAKTYTALSKVAPLPVVAEVMFFNVAAFGLA
LAWLTTWATSGLAGRRIWDAALVAASPLVIFQIFTNFDALATGLATSGLLAWARRRPVLAGVLIGLG
SAAKLYPLLFLYPLLLLIRAGRLNALARTMAAAAATWLLVNLPMMLLPRGWSEFFRLNTRRGDDM
DSLYNVKSFTGWRGFDPTLGFWEPPVLNTVVTLLFVLCCAAIAYIALTAPHRPRVAQLTFLTVASFL
15 LVNKVWSPQFSLWLVLAVLALPHRRILLAWMTIDALVWVPRMYLYGNPSRSLPEQWFTTTVLLRD
IAVMVLCGLVWQIYRPGRDLVRTGGPGALPACGGVDDPVGGVFANAADAPPGRLP SWLRPRLGD
EHARERTPDAGRDRTFSGQHRA

>Rv0106 - TB.seq 124372:125565 MW:43701 SEQ ID NO:160
20 MRTPVILVAGQDHTDEVTGALLRRTGTVVVEHRFDGHVRRMTATLSRGELITTEDALEFAHGCVSC
TIRDDLLVLLRRLHRRDNVGRIVVHLAPWLEPQPICWAIDHVRVCVGHGYPDGPAALDVRVAAVVTC
VDCVRWLPQSLGEDELPDGR TVAQVTVGQAEFADLLVLTHPEPVAVAVLRR LAPRARITGGVDRVEL
ALAHLDNDSRRGRDTDPHTPLLAGLPPLAADGEVAIVEFSARRPFHPQRLHAAVDLLLDGVRTRGR
LWLANRPDQVMWLESAGGGLRVASAGKWLAAAMAASEVAYVDLERRLFADLMWVYPFGDRHTAMT
25 VLVCADPTDIVNALNAALLSDDEMASPQRWQSYVDPFGDWHDDPCHEMPDAAGEFSAHRNSGES
R

>Rv0125 - TB.seq 151146:152210 MW:34927 SEQ ID NO:161
MSNSRRRSLRWSWLLSVLAAVGLGLATAPAQAAPPALSQDRFADFPALPLDPSAMVAQVGPQVVNI
30 NTKLGYNNAVGAGTGIVIDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG YDRTQDVAVLQLR
GAGGLPSAAIGGGVAVGEPVAMGNSGGQGGTPRAVPGRVVALGQTVQASDSLGAETLNGLIQ
FDAAIQPGDSGGPVNGLGQVGMNTAASDNFQLSQGGQGFAPIGQAMAIAQGIRSGGGSPTVHI
GPTAFLGLGWDNNGNGARVQRVWGSAPAASLGISTGDVITAVDGAPINSATAMADALNGHHPGDVI
SVTWQTKSGGTRTGNVTLAEGPPA

35
>Rv0350 dnaK 70 kD heat shock protein, chromosome replication TB.seq 419833:421707
MW:66832 SEQ ID NO:162

MARAVGIDLGTNSVSVLEGGDPVVANSEGSRTTPSIVAFARNGEVLVGQPAKNQAVTNVDRTV
RSVKRHMGS DWSIEIDGKKYTAPEISARILMKLKRDAEAYLGEDITDAVITTPAYFNDAQRQATKDAG
QIAGLNVL RIVNEPTAAALAYGLDKGEKEQRILVFDLGGGTDFVSLLEIGEGWEVRATSGDNHLGGD
DWDQRRVVDWLVDKFKGTSGIDLT KD K MAMQRLREAAEKAKIELSSSQSTSINLPYITVDADKNPLFLD
5 EQLTRAEFQRITQDLLDRTRKPFQSVIADTGISVSEIDHVVVLVGGSTRMPAVTDLVKELTGGKEPNKG
VNPDEVVAVGAALQAGVLKGEVKDVL LLDVTPLSLGIETKGGVMTRLIERNTTIPTKRSETFTTADDN
QPSVQIQVYQGEREIAAHNKLLGSFELTGIPAPRGIPQIEVTFDIDANGIVHTAKDKGTGKENTIRIQ
EGSGLSKEDIDRMIKDAEAHAEEDRKRREEADVRNQAETLVYQTEKFVKEQREAEAGGSKVPEDTLN
KVDAAVAEAKAALGGSDISAISAMEKLGQESQALGQAIYEAQAASQATGAHPGGEPGGAHPGS
10 ADDVDAEVDDGREAK

>Rv0351 grpE stimulates DnaK ATPase activity TB.seq 421707:422411 MW:24501
SEQ ID NO:163

VTDGNQKPDGNSGEQVTVTDKRRIDPETGEVRHVPPGDMPGGTAAADAHAHTEDKVAELTADLQRV
15 QADFANYRKRALRDQAAADRAKASVVSQLLGVLDLDERARKHGDLESGPLKSVADKLDSALTGLG
LVAFGAEGEDFDPVLHEAVQHEGDGGQGSKPVIGTVMRQGYQLGEQVLRHALVGVDTVVDAE
LESVDDGTAVADTAENDQADQGNSADTSGEQAESEPSGS

>Rv0352 dnaJ acts with GrpE to stimulate DnaK ATPase TB.seq 422450:423634 MW:41346
20 SEQ ID NO:164

MAQREWVEKDFYQELGVSSDASPEEIKRAYRKLARDLHPDANPGNPAAGERFKAVSEAHNVLS DPA
KRKEYDETRRLFAGGGFGGRRFD SGFGGGFGGFGVGGDGAEFNLNDLFDAASRTGGTTIGDLFGG
LFGRGGSARPSRPRRGNDLETETELDFVEAAKGVAMPLRLTSPAPCTNCHGSGARPGTSPKVCPTC
NGSGVINRNQGAFGFSEPCTDCRGSGSIIHPCEECKGTGVTRTRTINVRIPPGVEDGQRI RLAGQ
25 GEAGLRGAPSGDLYVTVHVRPDKIFGRDGD DLT VTPVVSFTELALGSTLSVPTLDGT VGV RVPKGTA
DGRILRVGRGVPKRSGGSGDLLVTVKVAVPPNLAGAAQEALEAYAAAERSSGFNPRAGWAGNR

>Rv0363c fba fructose bisphosphate aldolase TB.seq 441266:442297 MW:36545
SEQ ID NO:165

MPIATPEVYAEMLGQAKQNSYAFFAINCTSSETVNAAIKGFADAGSDGIIQFSTGGAIEFGSGLGVKDM
30 VTGAVALAEFTHVIAAKYPVNVALHTDHCPKDKLDSYVRPLLAISAQRVSKGGNPLFQSHMW DGS AV
PIDENLAI AQELLKAAAAAKIIEIGVVGGEEDGVANEINEKLYTSPEDFEKTIEALGAGEHGKYLAA
TFGNVHGVYKPGNVKLRPDILAQQQQVAAAKLGLPADAKPFDVVFHGGSGSLKSEIEEALRYGVVKM
NVDTD TQYAFTRPIAGHMFTNYDGV LKVDGEVGVKKVYDPRSYLKAEASMSQRVVQACNDLHCA
GKSLTH

35 >Rv0405 pks6 TB.seq 485729:489934 MW:147615 SEQ ID NO:166

MTDGSVTADKLQKWFREYLS THIEHPNEVSLDVPIRD LGLKSIDVLAIPGDLGDRFGFCIPDLAVWD
NPSANDLIDSLNQRSADSLRESHGHADRNTQGRGSINEPVAVIGVGCRFP GDIDGPERLWDFLTEK

KCAITAYPDRGFTNAGTFAESGGFLKDVAGFDNRFFDIPPDEALRMDPQQRLLLEVSWEALEHAGIIP
ESLRLSRTGVFVGVSSSTDYVRLVSASAQQKSTIWDNTGGSSSIANRISYFLDIQGSPSIVDTACSSSLV
AVHLACRSLSTWDCDIALVGGTNVLISPEPWGGFREAGILSQTGCCCHAFDKSADGMVRGEGCGVIVL
QRLSDARLEGRRLAILTGSAVNQDGKSNIGMAPNPSAQIGVLENACKSARVDPLEIGYVEAHGTGTS
5 LGDRIEHALGMVFGRKRPGSGPLMIGSIKPNIGHLEGAAGIAGLIKAVLMVERGSLLPSGGFTEPNP
AIPFTELGLRVDELQEWPVVAGRPRRAGVSSFGFGGTNAHVIVEEAGSVGADTVSGRADVGGSGG
GYVAWVISGKTASALAAQAGRLGRYVRARPALDVVDVGYSLVSTRSVFDHRAVVGQTRDELLAGL
AGVWAGRPEAGVVCVGKPAKTAFFVAGQGSQWLGMGSELYAAYPVFAEALDAVDELDRHLRY
PLRDVIWGHDDLLNTTEFAQPALFAVEVALYRLLMSWGVVRPGLVLGHVSGELAAAHVAGALCLPD
10 AAMLVAARGRLMQALPAGGAMFAVQAREDEVAPMLGHDVSIAAVNGPASVVISGAHDAVSAIADRL
RGQGRRRVHRLAVSHAFHSALMEPMIAEFTAVAAELSVGLPTIPVISNVTGQLVADDFASADYWARHIR
AVVRFGDSVRSACAGASRFIEVGPGGGLTSLIEASLADAQIVSVPTLRKDRPEPVSVMTAAAQGFV
SGMGLDWASVFSGYRPKRVELPTYAFQHQQFWLAPAPSVSDPTAAGQIGASDGGAEGLASSGFAA
RLAGRSADEQLAAAEVVCHEAAAVLGRDGAAGLDAGQAFADSGFNLSAVELRNRLTAVTAVTLPA
15 TAIFDHPTPTELAQYLITQIDGHGSSAAAAANPAERIDALTDLFLQACDAGRADGWKMVALASNTRE
RMSSPVRNNVSKNVALLADGISDVVICIPTLTVLSDQREYRDIANAMTGRHSVYSLTLPGFDSSDAL
PQNADMIVETVSNAIIDVVGSGCRFVLSGYSSGGVLAYALCSHLSVKHQRNPLGVALIDTYLPSQIAN
PSMNEGFSPNDTGKGLSREVIRVARMLNRLTATRLTAAATYAAIFQAWEPGRSMAPVLNIVAKDRIAT
VENLREERINRWRTAAAEAAYSVAEVPGDHFGMMSTSSEAIATEIHDWISGLVRGPHR

>Rv0435c - ATPase of AAA-family TB.seq 522348:524531 MW:75315 SEQ ID NO:167

VTHPDPARQLTLTARLNTSAVDSRRGVRLHPNAIAALGIREWDAVSLTGSRTTAAVAGLAAADTAV
GTVLLDDVTLNAGLREGTEVIVSPVTVYGARSVTLSGSTLATQSVPPVTLRQALLGKVMVTGDAVSL
LPRDLGPGTSTSASRALAAVGVISWTSSELLTVTGVDPDGPVSVQPNSLVTWGAGVPAAMGTSTAG
25 QVSISSPEIQIEELKGAQPQAAKLTEWLKLALDEPHLLQTLGAGTNLGLVLSGPAGVGKATLVRAVCD
GRRLLVTLDGPEIGALAAGDRVKAVASAVQAVRHEGGVLLITDADALLPAAAEPVASLILSELRTAVATA
GVVLIATSARPDQLDARLRSPELCDRELGLPLPDAATRKSLLLEALLNPVPTGDLNLDEIASRTPGFVVA
DLAALVREAALRAASRASADGRPPMLHQDDLLGALTIVIRPLSRSASDEVTVGDVTLDDVGDMAAAK
QALTEAVLWPLQHPDTFARLGVPEPRGVLLYGPPGCGKTFVVRALASTGQLSVHAVKGSELMDKWV
30 GSSEKAVRELFRARDSAPSLVFLDELDAAPRRGQSFDGVSVDRAVVAALLTELGDIDPLRDVVMLG
ATNRPDLIDPALLRPGRRLERLVFVEPPDAAARREILRTAGKSIPLSSDVLDLDEVAAGLDGYSAADCVAL
LREAALTAMRRSIDAANVTAADLATARETVRASLDPLQVASLRKFGTKGDLRS

>Rv0436c pssA CDP-diacylglycerol-serine o-phosphatidyltransferase TB.seq 524531:525388

MW:31219 SEQ ID NO:168

MIGKPRGRRGVNLQILPSAMTVLSICAGLTAIKFALEHQPKAAMALIAAAAILDGLDGRVARILDAQSR
MGAEIDSLADAVNFGVTPALVLYVSMLSKWPVGWVVLLYAVCVLRLARYNALQDDGTQPAYAHE

FFVGMPAPAGAVSMIGLLALKMQFGEGWWTSGWFLSFVWTGTSILLVSGIPMKMHAVSVPPNYAA
ALLAVLAICAAAVALAPYLLIWWIIIAYMCHIPFAVRSQRWLAQHPEVWDDKPKQRRRAVRRASRRHP
YRPSMARLGLRKPGRRRL

5 >Rv0440 groEL 260 kD chaperonin 2 TB.seq 528606:530225 MW:56728 SEQ ID NO:169
MAKTIAYDEEARRGLERGLNALADAVKVTLGPKGRNVVLEKKWGAPTITNDGVSLAKEIELEDPYEKI
GAELVKEVAKKTDDVAGDGTATVLAQALVREGLRNVAAGANPLGLKRGIEKAVEKVTTLLKGAK
EVETKEQIAATAAISAGDQSIGDLIAEAMDKVGNVITVEESNTFGLQLELTEGMRFDKGYISGYFVT
DPERQEAVLEDPYILLVSSKVSTVKDLLPLEKVIGAGKPLIIAEDVEGEALSTLVNKGRTFKSVAVK
10 APGFGDRRKAMLQDMAILTGGQVISEEVGLTLENADLSLLGKARKVVTKDETTIVEGAGDTDAIAGR
VAQIRQEIENSDDSDYDREKLQERLAKLAGGVAVIKAGAATEVELKERKHRIEDAVRNAKAAVEEGIVA
GGGVTLQAAPTLDLDELKLEGDEATGANIVKVALEAPLKQIAFNSGLEPGVVAEKVRNLPAGHGLNAQT
GVYEDLLAAGVADPVKVTRSAQNAAASIAGLFLTTEAVVADKPEKEKASVPGGGDMGGMDF

15 >Rv0482 murB TB.seq 570537:571643 MW:38522 SEQ ID NO:170
MKRSGVGSFAGAHIAEAVPLAPLTTLRVGPIARRVITCTSAEQVVAALRHLSAAKTGADRPLVFAG
GSNLVIAENLTDLTVVRLANSGITIDGNLVRAEAGAVFDDVVVRAIEQGLGGLECLSGIPGSAGATPVQ
NVGAYGAEVSDTITRVLDRCTGEVRWVSARDLRFGYRTSVLKHADGLAVPTVVLEVEFALDPSGR
SAPLRYGELIAALNATSGERADPQAVREAVLALRARKGMVLDPTDHDTSVGSFFTNPVVTQDVYE
20 RLAGDAATRKGDPVPHYPAPDGVKLAAGWLVERAGFGKGYPDAGAAPCRLSTKHALALTNRGGAT
AEDVVTLARAVRDGVHDFGITLKPEPVIGCML

>Rv0483 - TB.seq 571708:573060 MW:47859 SEQ ID NO:171
VVIRVLFPRVSLIPVNNSSTPQSQGPISRRLALTALGFGVLAPNVLVACAGKVTKLAEKRPPAPRLTF
25 RPADSAADVPIAPISVEVGDFQFQVALTNSAGKVAGAYSRRDRTIYTITEPLGYDTTYTWSGSAV
GHDGKAVPVAGKFTTVAPVKTNAGFQLADGQTVGIAAPVIIQFDSPISDKAAVERALTVTTDPPVEGG
WAWLPDEAQQARVHWRPREYYPAGTTVDVDAKLYGLPFGDGAYGAQDMSLHFQIGRRQVVAEV
SSHRIQVVTDAGVIMDFPCSYGEADLARNVTRNGIHVVTEKYSDFYMSNPAAGYSHIHERWAVRISN
NGEFIHANPMSAGAQQNSNVTNGCINLSTENAEQYYRSAVYGDPEVTGSSIQLSYADGDIWDWAV
30 DWDTWVSMSALPPPAKPAATQIPVTAPVTPSDAPTPSGTPTTTNGPGG

>Rv0489 gpm phosphoglycerate mutase I TB.seq 578424:579170 MW:27217 SEQ ID NO:172
MANTGSLVLLRHGESDWNALNLTGWDVGLTDKGQAEAVRSGELIAEHDLLPDVLYTSLRRAITT
AHLALDSADRLWIPVRRSWRLNERHYGALQGLDKAETKARYGEEQFMAWRRSYDTPPPPIERGSQ
35 FSQDADPRYADIGGGPLTECLADVVARFLPYFTDVIVGDLRVGKTVLIVAHGNSLRALVKHLDQMSDD
EIVGLNIPTGIPLRYDLDSAMRPLVRGGTYLDPEAAAAAGAAVAGQGRG

>Rv0490 senX 3sensor histidine kinase TB.seq 579347:580576 MW:44794 SEQ ID NO:173
VTVFSALLLAGVLSALALAVGGAVGMRLTSRVVEQRQRVATEWSGITVSQMLQCIVTLMPLGAAVVD
THRDVYVLNERAKELGLVRDRQLDDQAWRAARQALGGEDVEFDLSPRKRSATGRSGLSVHGHARL
LSEEDRRFAVVFVHDQSDYARMEAARRDFVANVSHELKTPVGAMALLAEALLASADDSETVRRFAE
5 KVLIEANRLGDMVAELIELSRLQGAERLPNMTDVDVDIVSEISRHKVAADNADIEVRTDAPSNLRVL
GDQTLTALANLVSNAIAYSPRGSLSISRRRRGANIEIAVTDRGIGIAPEDQERVFERFFRGDKARS
RATGGSGGLGLAIVKHAANHDGTIRVWSKPGTGSTFTLALPALIEAYHDDERPEQAREPELRSNRSQ
REEELSR

10 >Rv0500 proC pyrroline-5-carboxylate reductase TB.seq 590081:590965 MW:30172
SEQ ID NO:174
MLFGMARIAIIGGGSIGEALLSGLLRAGRQVKDLVAERMPDRANYLAQTYSVLVTSAADAVENATFV
VVAVKPADVEPVADLANATAAAENDSAEQVFVTVAGITIAYFESKLPAGTPVVRAMPNAAALVGAG
VTALAKGRFVTPQQLEEVSA LFDAVGGVLTVPESQLDAVTAVSGSGPAYFFLLVEALVDAGVGVGLS
15 RQVATDLAAQTMAGSAAMLLERMEQDQGGANGELMGLRVDLTASRLRAAVTSPGGTTAAALRELE
RGGFRMAVDAAVQAAKSRSSEQLRITPE

>Rv0528 - TB.seq 618303:619889 MW:57132 SEQ ID NO:175
MWRSLTSMGTALVLLFLLALAAIPGALLPQRGLNAAKVDDYLAHPLIGPWDELQAFDVFSSFWFTA
20 IYVLLFVSLVGCLAPRTIEHARSLRATPVAAPRNLARLPKHAHARLAGEPAALAATITGRLRGWRSITR
QQGDSVEVSAEKGYLREFGNLVFHFALLGLLVAVAVGKLFGEYEGNVIVIA DGGPGFCSASPAAFDSF
RAGNTVDGTSLHPICVRVNNFQAHYLP SGQATSFAADIDYQADPATADLIANSWRPYRLQVNHPLRV
GGDRVYLQGHGYAPTFTVTFPDGGQTRTSTVQWRPDNPQTLLSAGVVRIDPPAGSYNPNPDERRKHQI
AIQGLLAPTEQLDGTLLSSRFALNAPAVAIYRGDTGLDSGRPQSLFTLDHRLIEQGRLVKEKRVNL
25 RAGQQVRIDQGPAAGTVVRFDGAVPFVNLQVSHDPGQSWVLVFAITMMAGLLVSLVRRRRRVWARI
TPTTAGTVNVELGGLTRTDNSGWGAEFERLTGRLLAGFEARSPDMAEAAAAGTGRDVD

>Rv0667 rpoB [beta] subunit of RNA polymerase TB.seq 759805:763320 MW:129220
SEQ ID NO:176
30 LADSRQSKTAASPSRSPQSSSNNSVPGAPNRVSFAKLREPLEVPGLLDVQTDSFEWLIGSPRWRE
SAAERGDVNPVGGLLEEVLYELSPIEDFSGSMSLSFSDFPRFDDVKAPVDECKDKDMTYAAPLFVTAEF
INNNTGEIKSQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRS PGVYFDETIDKSTDKTLHSV KVIPSR
GAWLEFDVDKRDVTGVRIDRKRRQPVTLLKALGWTSEQIVERFGFSEIMRSTLEKDNTVGTDEALL
DIYRKLRPGEPPTKESAQTLLLENLFFKEKRYDLARVGRYKVNKKLGLHVGEPISSSTLTEEDVVATIEY
35 LVRLHEGQTTMTVPGGVEVPVETDDIDHFGNRRRLRTVGELIQNQIRVGMSRMERVVRERMTTQDVE
AITPQTLINIRPVAAIKEFFGTSQLSQFMDQNNPLSGLTHKRRLSALGPGGLSRERAGLEVRDVHPS
HYGRMCPIETPEGPNIGLIGLSVYARVNPFGFIETPYRKVVDGVVSDEIVYLTAEEDRHVVAQANS

PIDADGRFVEPRVLVRRKAGEVEYVPSSEVDYMDVSPRQMVSVATAMIPFLEHDDANRALMGANMQ
 RQAVPLVRSEAPLVGTGMELRAAIDAGDVVAEESGVIEVSADYITVMHONGTRRTYRMRKFARSN
 HGTCANQCPIVDAGDRVEAGQVIADGPCTDDGEMALGKNLLVAIMPWEGHNYEDAILSNRLVEEDV
 LSIHIEEHEIDARDTKLGAEETRDIPNISDEVLADLDERGIVRIGAEVRDGDILVGKVTPKGETELTPE
 5 ERLLRAIFGEKAREVRDTSCLKVPHGESGKVGIRVFSREDEDELPAQVNLVRVYVAQKRKISDGDKL
 AGRHGNKGVIGKILPVEDMPFLADGTPVDIILNTHGVPRRMNIGQILETHLGWCAHSGWKVDAAGV
 PDWAARLPDELLEAQPNAIVSTPVFDGAQAEALQGLLSCTLPNRDGDVLDADGKAMLFDGSRGEP
 FPYPVTVGMYIMKLHHLVDDKIARSTGPYSMITQQPLGGKAQFGGQRFGECEWAMQAYGAAY
 TLQELLTIKSDDTVGRVKVYEAVKGENIPEPGIPESFKVLLKELQSLCLNVEVLSSDGAAIELREGEDE
 10 DLERAAANLGINLSRNESASVEDLA

>Rv0668 rpoC [beta]' subunit of RNA polymerase TB.seq 763368:767315 MW:146740

SEQ ID NO:177

VLDVNFDELRIGLATAEDIRQWSYGEVKKPETINYRTLKPEKDGLFCEKIFGPTRDWECYCGKYKRV
 15 RFKGIICERCGVEVTRAKVRRERMGHIELAAPVTHIWYFKGVPSRLGYLLDLAPKDLEKIIYFAAYVITS
 VDEEMRHNELSTLEAEMAVERKAVEDQRDGELEARAQKLEADLAELEAEGAKADARRKVRDGER
 EMRQIRDRAQRELDRLIEDIWSTFTKLAPKQLIVDENLYRELVDRYGEYFTGAMGAESIQKLIENFDIDA
 EAESLRDVIRNGKGQKLRALKRLKVAAFQQSGNSPMGMVLDVAVPIPELRPMVQLDGGFRFATS
 DLNDLYRRVINRNNRLKRLIDLGAPEIIVNNEKRMLQESVDALFDNGRRGRPVTGPGNRPLKSLSDLL
 20 KGKQGRFRQNLGKRVDSGRSVIVGVPQLKLHQCGLPKLMALELFKPFVMKRLVDLNHAQNIKSAK
 RMVERQRPQVWDVLEEIVAEHPVLLNRAPTLHRLGIQAFEPMLVEGKAIQLHPLVCEAFNADFDGDQ
 MAVHLPLSAEAQAEARILMLSSNNILSPASGRPLAMPRLDMVTGLYYLTTEVPGDTGEYQPASGDHP
 ETGVYSSPAEAIMAADRGVLSVRKIKVRLTQLRPPVEIEAEFGHSGWQPGDAWMAETTLGRVMF
 NELLPLGYPFVNKQMHKKVQAAIINDLAERYPMIVVAQTVDKLDAGFYWATRSGVTVMADVLPVP
 25 RKKEILDHYEERADKVEKQFQRGALNHDERNEALVEIWKEATDEVGQALREHYPDDNPIITIVDSGAT
 GNFTQTRTLAGMKGLVTNPKGEFIPRPVKSSFREGLTVLEYFINTHGARKGLADTALRTADSGYLTRR
 LVDVSQDVIVREHDCQTERGIVVELAERAPDGTLRDPYIETSAYARTLGTDVDEAGNVIVERGQDL
 GDPEIDALLAAGITQVKVRSVLTCTSTGVCATCYGRSMATGKLDIGEAVGIVAAQSIGEPGTQLTM
 RTFHQGGVGEDITGGLPRVQELFEARVPRGKAPIADVTGRVRLEDGERFYKITIVPDDGGEEVYDKI
 30 SKRQRLRVFKHEDGSERVLSGDHVEVGQQLMEGSADPHEVLRVQGPREVQIHLVREVQEYVRAQ
 GVSIDHKHIEVIVRQMLRRVTIIDSGSTEFLPGSLIDRAFEAENRRVVAEGGEPAAGRPLMGITKAS
 LATDSWLSAASFQETTRVLTDAAINCRSDKLNGLKENVIIGKLIPAGTGINRYRNIAVQPTEEARAAAYT
 IPSYEDQYYSPLDGAATGAAPLDDYGYSDYR

35 >Rv0711 atsA TB.seq 806333:808693 MW:86216 SEQ ID NO:178

MAPEATEAFNGTIELDIRDSEPDPWGPYAAPVAPEHSPNLYLVWDDVGIATWDCFGGLVEMPAMTRV
 AERGVRLSQFHHTALCSPTRASLLTGRNATTVGMATIEEFTDGFPNCNGRIPADTALLPEVLAEHGYN

TYCVGKWHLTPLEESNMASTKRHWPTSRGFERYGFLGGETDQWYPDLVYDNHPVSPPGTPEGG
YHLSKDIADKTIEFIRDAKVIAPDKPWFSYVCPGAGHAPHHVFEWADRYAGRFDMGYERYREIVLE
RQKALGIVPPDTELSPINPYLDVPGPNGETWPLQDTPRPWDSLSDEEKKLFCRMAEVFAGFLSYTDA
QIGRILDYLEESGQLDNTIIVISDNGASGEGGPNNGSVNEGKFFNGYIDTVAESMKLFDHLGGPQTYN
5 HYPIGWAMAFNTPYKLFKRYASHEGGIADPAIISWPNGIAAHGEIRDNYVNVSDITPTVYDLLGMTTP
GTVKGIPQKPMDGVSFIAALADPAADTGKTTQFYTMLGTRGIWHEGWFANTIHAATPAGWSNFNAD
RWELFHIAADRSQCHDLAAEHPDKLEELKALWFSEAAKYNGPLADLNLETMTRSRYLVSERASY
VYYPDCADVIGIGAAVEIRGRSFAVLADVTIDTTGAEGVLFKHGGAGHGHVLFVRDGRHLHYVYNFLGE
RQQLVSSSGPVPSGRHLLGVRYLRTGTVPNSTHTPVGDLELFFDENLVGALTNLTHPGTFGLAGAAI
10 SVGRNGGSASVSSHYEAPFAFTGGTITQVTVDVSGRPFEDVESDLALAFSRD

>Rv0764c - lanosterol 14-demethylase cytochrome P450 TB.seq 856683:858035 MW:50879

SEQ ID NO:179

MSAVALPRVSGGHDEHGHLEEFRTDPIGLMQRVRDECGDVGTFLAGKQVLLSGSHANEFFFR
15 GDDDLQAKAYPFMTPIFGEGVVFDA SPERRKEMLHNAALRGEQMKGHAATIEDQVRRMIADWGE
AGEIDLLDFFAELTIYTSSACLIGKKFRDQLDGRFAKLYHELERGTDPYVDPYLPPIESFRRRDEARN
GLVALVADIMNGRIANPPTDKSDRDMLDVLI AVKAETGTPRFSADEITGMFISMMFAGHHTSSGTASW
TLIELMRHRDAYAAVIDELDELYGGRSVSFHALRQIPQLENVLKETLRLHPPLIILMRVAKGEFEVQG
HRIHEGDLVAASPAISNRIPEDFDPHDFVPARYEQPRQEDLLNRWTWIPFGAGRHRVCVGAFAIMQI
20 KAIFSVLLREYEFEMAQPPESYRNDHSMVQLAQAPACVRYRRRTGV

>Rv0861c - DNA helicase TB.seq 958524:960149 MW:59773 SEQ ID NO:180

VQSDKTVLLEVDHELAGAARAAIAPFAELERAPEHVHTYRITPLALWNARAAGHDAEQVVDALVSYS
RYAVPQPLLVDIVDMARYGRLQLVKNPAHGLTLVSLDRAVLEEVLNRNKKIAPMLGARIDDDTVVHP
25 SERGRVKQLLLKIGWPAEDLAGYVDGEAHPISLHQEGWQLRDYQRLAADSFWAGGSGVWLPCGA
GKTLVGAAAMAKAGATTILVTNIVAARQWKRELVARTSLTENEIGEFSGERKEIRPVITISTYQMITRR
TKGEYRHLELFDSDRWGLIYDEVHLLPAPVFRMTADLQSKRRLGLTATLIREDGREGDVFSLIGPKR
YDAPWKDIEAQGWIAPAECVEVRVTMTDSERMMYATAEPEERYRICSTVHTKIAVVKSIKAKHPDEQ
TLVIGAYLDQLDELGAELGAPVIQGSTRTSREALFADFRRGEVATLVVSKVANFSIDLPEAAVAVQVS
30 GTFGSRQEEAQRLGRILRPKADGGGAIFYSVVARDSLDAEYAAHRQRFLAEQGYGYIIRDADDLLGP
AI

>Rv0904c accD3 TB.seq 1006694:1008178 MW:51741 SEQ ID NO:181

VSRIITDQLRHAVLDGRGSFVSWDSEPLAVPVADSYARELAAARAATGADESQVTGEGRVFGRRVAV
35 VACEFDLGGSIGVAAAERITA AVERATAERLPLLASPSSGGTRMQEGTVAFQMVKIAAAIQLHNQA
RLPYLVYLRHPTTGGVFASWGLGHLTVAEPGALIGFLGPRVYELLYGDPFPGSVQTAENLRRHGIID
GVVALDRLRPMLDRALTVLIDAPEPLPAPQTPAPVPDVPTWDSVVASRRPDRPGVRQLLRHGATDR

VLLSGTDQGEAATLLALARFGGQPTVVLGQQRAVGGGGSTVGPAAALREARRGMALAAELCLPLVL
VIDAAGPALSAAAEQGGLAGQIAHCLAELVLTDTPTVSILLGQSGGPALAMLPADRVLAALHGWLAP
LPPEGASAIVFRDTAHAAELAAAQGIRSADLLKSGIVDTIVPEYDAADEPIEFALRLSNAIAAEVHALR
KIPAPERLATRLQRYRRIGLPRD

5

>Rv0983 - TB.seq 1099064:1100455 MW:46454 SEQ ID NO:182

MAKLARVVGLVQEEQPSDMTNHPRYSPPPQQPGTPGYAQQGQQQTYSQQFDWRYPPSPPPQPTQY
RQPYEALGGTRPGLIPGVIPTMTPPPGMVQRQRPRAGMLAIGAVTIAVVSAGIGGAAASLVGFNRAPA
GPSGGPVAASAAPSIPAANMPPGSVEQVAAKVVPVVMLETDLGRQSEEGSGIILSAEGLILTNNHVI
10 AAAKPPLGSPPPKTTVTFSDGRTAPFTVVGADPTSDIAVVRVQGVSGLTPISLGSSSDLRVGQPVL
IGSPLGLEGTVTTGIVSALNRPVSTTGEAGNQNTVLDIAIQTDAAINPGNSGGALVNMQALVGVNSAI
ATLGADSADAQSGSIGLGFAPVDQAKRIADELISTGKASHASLGQVTNDKDTLGAKIVEVWAGGAA
ANAGVPKGVVTKVDDRPINSADALVAAVRSKAPGATVALTFQDPSGGSSRTVQVTLGKAEQ

15

>Rv1008 - Similar to E.coli protein YcfH TB.seq 1127087:1127878 MW:29066 SEQ ID NO:183

LVDATHLDACGARDADTVRSLVERAAAAGVTAVTVADDLESARWVTRAAEWDRRVYAAVALHPT
RADALDAAARAELERLVAHPRVVAVGETGIDMYWPGRLDGCAEPHVQREAFAWHIDLAKRTGKPLM
IHNQADRDVLDVLAEGAPDTVILHCFSSDAAMARTCVDAGWLLSLSGTVSFRTARELREAVPLMP
VEQLLVETDAPYLTPHPHRLANEPYCLPYTVRALAELVNRPEEVALITTSNARRAYGLGWMRQ

20

>Rv1009 - lipoprotein, similar to various other MTB proteins TB.seq 1128089:1129174 MW:38079
SEQ ID NO:184

MLRLVVGALLLVAFAGGYAVAACKTVTLTVDGTAMRVTTMKSRVIDIVEENGFSVDDRDDLYPAAG
VQVHDADTIVLRRSRPLQISLDGHDQVWTTASTVDEALAQLAMTDTAPAAASRASRVPLSGMALP
25 VSAKTVQLNDGGLVRTVHLPAPNVAGLLSAAGVPLLQSDHVVPAATAPIVEGMQIQVTRNRIKKVTE
RLPLPPNARRVEDPEMNMSREVVEDPGVPGTQDVTFAVAEVNGVETGRLPVANVVVTPAHEAVR
VGTKPGTEVPPVIDGSIWDAIAGCEAGGNWAINTGNGYYGGVQFDQGTWEANGGLRYAPRADLAT
REEQIAVAEVTRLRQGWGAWPVCAARAGAR

30

>Rv1010 ksgA 16S rRNA dimethyltransferase TB.seq 1129150:1130100 MW:34647
SEQ ID NO:185

MCCTSGCALTIRLLGRTEIRRLAKELDFRPRKSLGQNFVHDANTVRRVVAASGVSRSDLVLEVGPGL
GSLTLALLDRGATVTAVEIDPLLASRLQQTVAEHSSEVHRLTVVNRDVLALRREDLAAAPTAVVANL
PYNVAVPALLHLLVEFPSIRVVTVMVQAEVAERLAAEPGSKEYGVPSVKLRFFGRVRRCGMVSPVTF
35 WPIPRVYSGLVRIIDRYETSPWPTDDAFRRRVFELVDIAFAQRRKTSRNAFVQWAGSGSESANRLAA
SIDPARRGETLSIDDFVRLRRSGGSDEATSTGRDARAPDISGHASAS

>Rv1011 - Similar to E.coli protein YcbH TB.seq 1130189:1131106 MW:31350

SEQ ID NO:186

VPTGSVTVRPVKVNLVAVGDRREDGYHELTTFHAVSLVDEVTVRNADVLSLELVGEGADQLPTD
ERNLAWQAAELMAEHVGRAPDVSIMIDKSIPVAGGMAGGSADAAVLVAMNSLWELNVPRDLRML
5 AARLGSDVPFALHGGTALGTGRGEELATVLSRNTFHWVLAFADESGLLTSAVYNELDRLREVGDPRL
GEPGPVLAALAAGDPDQLAPLLGNEMQAAAVSLDPALARALRAGVEAGALAGIVSGSGPTCAFLCTS
ASSAIDVGAQLSGAGVCRTVRVATGPVPGARVVSAPTEV

>Rv1106c - cholesterol dehydrogenase TB.seq 1232845:1233954 MW:40743 SEQ ID NO:187

10 MLRRMGDASLTTELGRVLVTGGAGFVGANLVTLLDRGHWVRSFDRAPSLPAHPQLEVLQGDITD
ADVCAAVDGIDTIFHTAAIILMGGASVTDEYRQRSFAVNVGGTENLLHAGQRAGVQRFVYTSSNS
VVMGGQNIAGGDETLPTDRFNDLYTETKVVAERFVLAQNGVDGMLTCAIRPSGIWNGDQTMFRK
LFESVLKGHVKVLVGRKSARLDNSYVHNLIHGFIILAAHLVPDGTAPGQAYFINDAEPINMFEFARPV
EACGQRWPKMRISGPAVRWVMTGWQRLHFRFGFPAPLLEPLAVERLYLDNYFSIAKARRDLGYEPL
15 FTTQALTECLPYVSLFEQMKNEARAECTAATVKP

>Rv1110 lytB2 TB.seq 1236183:1237187 MW:36298 SEQ ID NO:188

MVPTVDMGIPGASVSSRSVADRPNRKRVLAEPRGYCAGVDRAVETVERALQKHGPPVYVRHEIVH
NRHVVDLAKAGAVFVEETEQVPEGAIVFSAHGVAPT VHVSASERNLQVIDATCPLVTKVHNEARR
20 FARDDYDILLIGHEGHEEVGTAGEAPDHVQLVDGVDQVTVRDEDKVWLSQTLSVDETMEIV
GRLRRRFPKLQDPPSDDICYATQNRQVAVKAMAPECELVIWVGSRNSSNSVRLVEVALGAGARAAH
LVDWADDIDSAWLDGVTTVGVTS GASVPEVLVRGVLERLAECGYDIVQPVTANETLVFALPRELRS
PR

25 >Rv1216c - TB.seq 1359473:1360144 MW:24863 SEQ ID NO:189

MHIGLKIFIWGLVGLVFGALLFGPAGTFDYWQAWVFLAAVSTTIGPTIYLARNDPAALQRRMRSGP
LAEGRTIQKFIVIGAFGLFFAMMVL SACDHRYGWSSVPAAVCVIGDVLVMTGLGIAMLVVIQNRYAAS
TVRVEAGQILASDGLYKIVRHPMYAGNVMMMTGIPLALGSYWAMFILVPGLVLVFRILDEEKLLTQEL
SGYREYRQLVRYRLVPYVW

30 >Rv1223 htrA TB.seq 1365810:1367456 MW:56547 SEQ ID NO:190

VSHLSQRMAGLLRVHGEWSRSVDTRVDTDNAMPARFSAQIQNEDEVTS DQGNGGPNGGGRLAP
RPVFRPPVDPASRQAFGRPSGVQGSFVAERVRPQKYQDQSDFTPNDQLADPVLQEA FGRPFAGAE
SLQRHPIDAGALAAEKDGAGPDEPD PWRDPAALAAALGTPALAAPAPHGALAGSGKLGVRDVLFGG
35 KVSYLALGILVAIALVIGGIGGVIGRKTAEVDAFTTSKVTLSTTGNAQEPAGRFTKVA AAVADSVVTIE
SVSDQEGMQGSGVIVDGRGYIVTNNH VISEAANNPSQFKITTVFNDGKEVPANLVGRDPKTDLAVLK
VDNVDNLTVARLGDSSKVRVGDEV LAVGAPLGLRSTVTQGIVSALHRPVPLSGEGSDTDTVIDAIQTD

ASINHGNSGGPLIDMDAQVIGINTAGKSLSDSASGLGFAIPVNEMKLVANSLIKDGKIVHPTLGISTRV
SNAIASGAQVANVKAGSPAQKGGILENDVIVKVGNNRAVADSDEFVAVRQLAIGQDAPIEVVREGRH
VTLTVKPPDST

5 >Rv1224 - TB.seq 1367461:1367853 MW:14083 SEQ ID NO:191

VFANIGWWEMLVLMVGLVVLGPERLPGAIRWAASALRQARDYLSGVTSQLREDIGPEFDDLRGHL
GELQKLRGMTPRALTKHLLDGDDSLFTGDFDRPTPKKPDAGSAGPDATEQIGAGPIPFDSAT

10 >Rv1229c mrp similar to MRP/NBP35 ATP-binding proteins TB.seq 1371778:1372947 MW:41064
SEQ ID NO:192

MPSRLHSAVMGSTRDGDNLAAIRTALGKVIDPELRRPITELGMVKSIDTGPDGSVHVEIYLTIAAGCPKK
SEITERVTRAVADVPGTSAVRVSLDVMSDEQRTELKQLRGDTREPVIPFAQPDSLTRYAVASGKG
GVGKSTVTNLAAMAVRGLSIGVLDADIHGHSIPRMMGTDRPTQVESMILPPIAHQVKVISIAQFTQ
GNTPVVWRGPMHLRALQQFLADVWGDLDVLLDLPDGTGDAISVAQLIPNAELLVVTTPQLAAAE
15 VAERAGSIALQTRQRIVGVVENMSGTLTDPGTTMQVFGEGGGRLVAERLSRAVGADVPLLQGIPLDP
ALVAAGDSGVPLVLSSPDSAIGKELHSIADGLSTRRRGLAGMSLGLDPTRR

>Rv1239c corA magnesium and cobalt transport protein TB.seq 1381943:1383040 MW:41470
SEQ ID NO:193

20 VFPGFDALPEVLRPVARPQPPNAHPVAQPPAQALVDCGVYVCGQRLPGKYTYAAALREVREIELTG
QEAFVWIGLHEPDENQMVDVDFGLHPLAVEDAVHAHQRPKLERYDETLFLVLKTVNYVPHEVSV
LAREIVKTGEIMIFVGKDFVTVRHGEHGLSEVRKRMDADPEHLRLGPYAVMHAIADYVVDHYLEV
NLMETDIDSIEEVAFAFGRKLDIEPIYLLKREVELRRCVNPLSTAFQRMQTESKDLISKEVRRYL
ADHQTEAADQIASYDDMLNSLVQAALARVGMQQNMDMRKISAWAGIIAVPTMIAGIYGMNFHFMPEL
25 DSRWGYPTVIGGMVLICFLYHVFRRNRNL

>Rv1279 - TB.seq 1430060:1431643 MW:57332 SEQ ID NO:194

MDTQSDYVVGTSAGAVVASRLSTDPATTVALEAGPRDKNRFIGVPAAFSKLFRSEIDWDYLTEP
QPELDGREIYWPRGKVLGGSSSMNAMMWVRGFASDYDEWAARAGPRWSYADVLGYFRRIENVTA
AWHFVSGDDSGVTGPLHISRQSPRSVTAAWLAAARECGFAAARPNSPRPEGFCETVVTQRRGAR
30 FSTADAYLKPMARRKNLRLVTGATATRVVIDGDRAVGVEYQSDGQTRIVYARREVLCAAGVNSPQL
LMLSGIGDRDHLAEHDIDTVYHAPEVGCNLLDHLVTVLGFDEKDSLFAAEKPGQLISYLLRRRGMLT
SNVGEAYGFVRSRPELKLPLELIFAPAPFYDEALVPPAGHGVVFGPILVAPQSRGQITLRSADPHAK
PVIEPRYLSDLGGVDRAAMMAGLRICARIAQARPLRDLLGSIARPRNSTELDEATLELALATCSHTLYH
PMGTCRMGSDEASVVDPQLRVRGVDGLRVADASVMPSTVRGHTHAPSVLIGEKAADLIRS

35 >Rv1294 thrA homoserine dehydrogenase TB.seq 1449373:1450695 MW:45522 SEQ ID NO:195

VPGDEKPGVAVLGLGNVGSEVVRIENSAEDLAARVGAPLVLRGIGVRRVTTDRGVPIELLTDDIEEL
VAREDDIVVEVMGPVEPSRKAILGALERGKSVVTANKALLATSTGELAQAAESAHDLYFEAAVAGA

IPVIRPLTQSLAGDTVLRVAGIVNGTTNYILSAMDSTGADYASALADASALGYAEADPTADVEGYDAA
AKAAILASIAFHTRVTADDVYREGITKVTPADFGSAHALGCTIKLLSICERITTTDEGSQRVSARVYPALV
PLSHPLAAVNGAFNAVVEAEAGRLMFYQGQAGGAPTASAVTGDLVMAARNRVLGSRGPRESKY
AQLPVAPMGFIETRYVSMNVADKPGVLSAVAAEFAKREVSIAEVRQEGVWDEGGRRVGARIVVWTH
5 LATDAALSETVDALDDLDVWQGVSSVIRLEGTGL

>Rv1323 fadA4 acetyl-CoA C-acetyltransferase (aka thil) TB.seq 1485860:1487026 MW:40049
SEQ ID NO:196

VIVAGARTPIGKLMGSLKDFSASELGAIKGALEKANVPASLVEYVIMGQVLTAGAGQMPARQAAVA
10 AGIGWDVPALTINKMCLSGIDAIALADQLIRAREFDVWVAGGQESMTKAPHLLMNSRSGYKYGDVTVL
DHMAYDGLHDVFTDQPMGALTEQRNDVDMFTRSEQDEYAAASHQKAAAAWKDGVFADDEVIPVNI
QRTGDPLQFTEDEGIRANTTAAALAGLKPAFRGDGTITAGSASQISDGAAAVVMNQEKAEQELGLTW
LAEIGAHHGVVAGPDSTLQSQPANAINKALDREGISVDQLDVVEINEAFAAVALASIRELGLNPQIVNVN
GGAIAVGHPLGMSGTRITLHAALQLARRGSGVGVAALCGAGGQGDALILRAG

>Rv1389 gmk putative guanylate kinase TB.seq 1564399:1565022 MW:22064 SEQ ID NO:197
VSVGEGPDTKPTARGQPAVGRVVLGSPSAVGKSTVVRCLRERIPNLHFSVSATTRAPRPGVEVDG
VDYHFIDPTRFQQLIDQGELLEWAEIHGGLHRSGLTAQPVRAAAATGVPVLIEVDLAGARAIKKTMP
15 AVTVFLAPPSWQDLQARLIGRGTTETADVIQRRLDTARIELAAQGDFDKVVVNRRLSACAEVLSLLVG
TAPGSP

>Rv1407 fmu similar to Fmu protein TB.seq 1583099:1584469 MW:48494 SEQ ID NO:198
MTPRSRGP RRRPLDPARRAAFETLRAVSARDAYANLVL PALLAQRGIGGRDAAFATELTYGTCRAR
GLLDAVIGAAAERSPQAIDPVLLDLLRLGT YQLLRTRVDAHA AVSTTVEQAGIEFDSARAGFVNGVLR
25 TIAGRDERSWVGELAPDAQNDPIGHAA FVHAHPRWIAQAFADALGA AVGELEAVLASDDERPAVHLA
ARPGVLTAGELARAVRGTVGRYSPFAVYLPRGDPGR LAPVRDGGQALVQDEGSQLVARALT LAPVDG
DTGRWLDLCAGPGGKTALLAGLGLQCAARVTAVEPSPHRADLVAQNTRGLPVELLRVDGRHTDLDP
GFDRVLVDAPCTGLGALRRRPEARWRRQPADVAALAKLQRELLSAAIALTRPGGVVLYATCSPHLAE
TVGAVADALRRHPVHALDTRPLFEPVIAGLGEGPHVQLWPHRHGTDAMFAAALRRLT

>Rv1409 ribG riboflavin biosynthesis TB.seq 1585192:1586208 MW:35367 SEQ ID NO:199
MNVEQVKSIDEAMGLAIEHSYQVKGTTYPKPPVGAVIVDPNGRIVGAGGTEPAGGDHAEVVALRRAG
GLAAGAIWVVTMEPCNHYGKTPPCVNALIEARVGT VVYAVADPNGIAGGGAGRLSAAGLQVRSGVLA
EQVAAGPLREWLHKQRTGLPHVTWKYATSIDGRSAAADGSSQWISSEARLDLHRRRAIADAILVGT
35 GTVLADDPALTARLADGSLAPQQPLRVVVGKRDIPPEARVLNDEARTMMIRTHEPMEVLRALSDRTD
VLLEGGPTLAGAFLRAGAINRILAYVAPILLGGPVTA VDDVGVS NITNALRWQFDSVEKVGPDLLLSLV
AR

>Rv1440 secG TB.seq 1617715:1618065 MW:12140 SEQ ID NO:200

VAGVTAAVSARLKADEARRPGFYAAGSGPLPQVRGSTLPVMELALQITLIVTSVLVLLVLLHRAKGG
GLSTLFGGGVQSSLSGSTVVEKNLDRLTLFVTGIWLVSIGVALLIKYR

5 >Rv1484 InhA TB.seq 1674200:1675006 MW:28529 SEQ ID NO:201

MTGLLDGKRILVSGIITDSSIAFHARVAQEQAQLVLTGFDRLRLIQRITDRLPAKAPLLELDVQNEEH
LASLAGRVTEAIGAGNKLDGVVHSIGFMPQTGMGINPFFDAPYADVSKGIHISAYSASYASMAKALLPIM
NPGGSIVGMDFDPSRAMPAYNWMTVAKSALESVNRFVAREAGKYGVRNLSVAAGPIRTLAMSAIVG
GALGEEAGAQIQLLEEGWDQRAPIGWNMKDATPVAKTVCALLSDWLPATTGDIIYADGGAHTQLL

10

>Rv1617 pykA pyruvate kinase TB.seq 1816187:1817602 MW:50668 SEQ ID NO:202

VTRRGKIVCTLGPATQRDDLVRALVEAGMDVARMNFSGDYDDHKVAYERVRVASDATGRAVGVL
ADLQGPKIRLGRFASGATHWAEGETVRITVGACEGSHDRVSTTYKRLAQDAVAGDRVLVDDGKVAL
WDAVEGDDWCTVVEGGPVSDNKGISLPGMNVTA PALSEKDIEDLTFALNLGVD MVALSFVRSPAD
15 VELVHEVMDRIGRRVPVIAKLEKPEAIDNLEAIVLAFDAVMVARGDLGVELPLEEVPLVQKRAIQMARE
NAKPVIVATQMLDSMIENS RPTRA EASDVANAVLDGADALMLSGETSVGKYPLAAVRTMSRIICAVEE
NSTAAPPLTHIPRTKRGVISYAARDIGERLDAKALVAFTQSGD TVRRLARLHTPLPLLAFTAWPEVRS
QLAMTWGTETFIVPKMQSTDGMIRQVDKS LLELARYKRGDLVVIVAGAPPGTVGSTNLHVHRIGEDD
V

20

>Rv1630 rpsA 30S ribosomal protein S1 TB.seq 1833540:1834982 MW:53203 SEQ ID NO:203

MPSPTVTSPQVAVNDIGSSEDFLAAIDKTIKYFNDGDIVEGTIVKVDRDEVLLDIGYKTEGVIPARELSIK
HDVDPNEVSVSGDEVEALVLTKE DKEGRILSKKRAQYERAWGTIEALKEKDEAVKGTVIEVVKGLI
LDIGLRGFLPASLVEMRRVRDLQPYIGKEIEAKIIELDKNRNNVLSRRRAWLEQTQSEVRSEFLNNLQK
25 GTIRKGVVSSIVNFGAFVDLGGVDGLVHVSELSWKHIDHPSEVVQVGDEVTVEVLDVDMDRERVSL
LKATQEDPWRHFARHTAIGQIVPGKVTKLVPFGA FVRVEEGIEGLVHISELAERHVEVPDQVVAVGDD
AMVKVIDIDLERRRISLSLKQANEDYTEEFDAKYGMADSYDEQGN YIFPEGFDAETNEWLEGFEKQ
RAWEARYAEAERRHKMHTAQMEKFAAAEAAGRGADDQSSASSAPSEKTAGGSLASDAQLAALRE
KLAGSA

30

>Rv1631 - TB.seq 1835011:1836231 MW:44669 SEQ ID NO:204

MLRIGLTGGIGAGKSLLSTTFSQCGGIVDGDVLAREVVQPGTEGLASLVDAFGRDILLADGALDRQA
LAAKA FRDDES RGV L NGIVHPLVARRRSEIIAAVSGDAVVVEDIPLLVESGMAPLFPLVVVVHADVELR
VRRLVEQRGMAEADARARIAAQASDQRRRAVADVWLDNSGSPEDLVRRARDVWNTRVQPFANL
35 AQRQIARAPARLPADPSWPDQARRIVNRLKIACGHKALRVDHIGSTAVSGFPDFLAKDVIDIQVTE
SLDVADELAEP LLAAGYPRLEHITQDTEKTDARSTVGRYDHTD SAALWHKRVHASADPGRPTNVHLR

VHGWPNQQFALLFVDWLAANPGAREDYLTVKCDADRRADGELARYVTAKEPWFLDAYQRAWEWA
DAVHWRP

>Rv1706c - TB.seq 1932695:1933876 MW:39779 SEQ ID NO:205

5 MTLDVPVNQGHVPPGSGVACCLVGVTAVADGIAGHSLSNFGALPPEINSGRMYSGPGSGPLMAAAAA
WDGLAAELSSAATGYGAAISELTNMRWWSGPASDSMVAAVLPFVGWLSTTATLAEQAAMQARAAA
AAFEAAAFAMTVPPPAIAANRTLLMTLVDTNWFGQNTPAIATTESQYAEMWAQDAAAMYGYASAAAP
ATVLTTPFAPPPQTTNATGLVGHATAVAALRGQHSWAAAIPWSDIQKYWMMFLGALATAEGFIYDSG
GLTLNALQFVGGMLWSTALAEAGAAEAAAGAGGAAGWSAWSQLGAGPVAASATLAAKIGPMSVPP
10 GWSAPPATPQAQTVARSIPGIRSAAEAAETSULLRGAPTGRSRAAHMGRRYGRRLTMADRPNVG

>Rv1745c - similar to Q46822 ORF_O182 TB.seq 1971381:1971989 MW:22490 SEQ ID NO:206

MTRSYRPAPPIERVLLNDRGDATGVADKATVHTGDTPLHLAFSSYVFDLHDQLLITRRAATKRTWP
AVWTNSCCGHPLPGESLPGAIRRRRLAAELGLTPDRVDLILPGFRYRAAMADGTVENEICPVYRVQVD
15 QQPRPNSDEVDAIRWLSWEQFVRDVTAGVIAPVSPWCRSQLGYLTKLGPCPAQWPVADDCRLPKA
AHGN

>Rv1800 - TB.seq 2039451:2041415 MW:67068 SEQ ID NO:207

MLPNFAVLPPVNSARVFAGAGSAPMLAAAAAWDDLASELHCAAMSFGSVTSGLVVGWWQGSASA
20 AMVDAAASYIGWLSTSAHAEGAAGLARAASVSFEEALAATVHPAMVAANRAQVASLVASNLFGQN
APAIAALESLEYECMWAQDAAAMAGYYVGASAVATQLASWLQRLQSIPGAASLDARLPSSAEAPMGV
VRAVNSAIAANAAAAQTVGLVMGSGTPIPSARYVELANALYMSGSVPGVIAQALFTPQGLYPVVVIK
NLTFDSSVAQGAVILESARQQIAAGNNVTVFGYSQSATISSLVMANLAASADPPSPDELSTLIGNPN
NPNGGVATRFPGISFPSLGVATGATPHNLYPTKIYTIEYDGVADFPRYPLNFVSTLNAIAGTYVHSN
25 YFILTPEQIDAAVPLTNTVGPTMTQYYIIRTENLPLEPLRSVPVIGNPLANLVQPNLKVIVNLGYGDPA
YGYSTSPPNVATPFGLFPEVSPVVIADALVAGTQQGIGDFAYDVSHLELPLPADGSTMPSTAPGSGT
PVPPLSIDSLIDDLQVANRNLANTISKVAATSYATVLPTADIANAALTIVPSYNIHLFLEGIQQALKGDPM
GLVNAVGYPLAADVALFTAAGGLQLLIISAGRTIANDISAIVP

30 >Rv1844c gnd 6-phosphogluconate dehydrogenase (Gram -) TB.seq 2093732:2095186
MW:51548 SEQ ID NO:208

MSSSESPAGIAQIGVTGLAVMGSNIAARNFARHGYTVAVHNRSVAKTDALLKEHSSDGKFSRSETIPEF
LAALEKPRRVLIMVKAGEATDADAVINELADAMEPGDIIIDGGNALTDTMRREKAMRERGLHFVGAG
ISGGEEGALNGPSIMPGGPAESYQSLGPLLEEISAHVDGVPCCTHIGPDGSGHFVKMVHNGIEYSDM
35 QLIGEAYQLMRDGLGLTAPAIADVTEWNNGDLSYLVEITAIEVLRQTDAKTGKPLVDVIVDRAEQKG
TGRWTVKSALDLGVPVTGIAEAVFARALSGSVGQRSAAASGLASGKLGEQPADPATFTEDVRQALYA
SKIVAYAQQFNQIQAGSAEFGWDITPGDLATIWRGGCIIRAKFLNHIKEAFDASPNLASLIVAPYFRGA

VESAIDSWRRVWSTAAQLGIPTPGFSSALSYYDALRTARLPAALTQAQRDFFGAHTYGRIDEPGKFHT
LWSSDRTEVPV

>Rv1900c lipJ TB.seq 2146246:2147631 MW:49685 SEQ ID NO:209

5 VAQAPHIHRTRYAKCGDMDIAYQVLGDGPTDLLVLPGFVPIDSIDDEPSLYRFHRRLASFSRVIRLDH
RGVGLSSRLAAITTLGPKFWAQDAIVMDAVGCEQATIFAPSFHAMNGLVLAADYPERVRS LIVNGS
ARPLWAPDYPVGAQVRRADPFLTVALEPDAVERGFDVLSIVAPT VAGDDVFRAWWDLAGNRAGPP
SIARAVSKVIAEADV RDVLGHIEAPT LILHRVGSTYIPVGHGRYLA EHIAGSRLVELPGTDTLYWVGDT
10 GPMLDEIEEFITGVRGGADAERMLATIMFTDIVGSTQHAAALGDDRWRDLDNHD TIVCHEIQRFGGR
EVNTAGDGFVATFTSPSAAIACADDIVDAVAALGIEVRIGIHAGEVEVRDASHGTDVAGVAVHIGARVC
ALAGPSEVLVSSTVRDIVAGSRHRFAERGEQELKGVPGRWRLCVMRDDATRTR

>Rv1967 - TB.seq 2210599:2211624 MW:36516 SEQ ID NO:210

15 MREN LGGVVRLGVFLAVCLLTAFLLI AVFGEVRF GDGKTYAEFANVSNLRTGKLVRIAGVEVGKVT
RISINPDATVRVQFTADNSVTLTRGTRAVIRYDNLFGDRYLAEEGAGGLAVLRPGHTIPLARTQPALD
LDALIGGFKPLFRALNPEQVNALSEQLLHAFAGQGPTIGSLLAQSAAVTN TLADRDRLIGQVITNLNV
LGSLGAHTDRLDQAVTSLSALIHLRAQRKTDISNAVAYTNAAAGSVADLLSQARAPLAKVVRETDRVA
GIAAADHDYLDNLLNTLPDKYQALVRQGMYGDFFAFYLCDVWLKVNGKGGQPVYIKLAGQDSGRCA
20 PK

>Rv1975 - TB.seq 2218050:2218712 MW:23650 SEQ ID NO:211

MSRRASATCALSATTAVAIMAAPAARADDKRLNDGVVANVYTVQRQAGCTNDVTINPQLQLAAQWH
TLDLLNNRHLNDDTGSDGSTPQDRAHAAGFRGKVAETVAINPAVAISGIELINQWYYNPAFFAIMSDC
ANTQIGVWSENSPDRTVVAVYQGPD RPSAMP PRGAVTGPPSPVAAQENVPIDSPDYDASDEIEY
25 GINWLPWILRGVYPPPPAMPPQ

>Rv1981c nrdF ribonucleotide reductase small subunit TB.seq 2224221:2225186 MW:36591
SEQ ID NO:212

30 MTGKLVERVHAINWNRLLDKDLQVWERLTGNFWLPEKIPLSNDLASWQTLSSTEQQTIRVFTGLT
LLDTAQATVGAVAMIDDAVTPHEEAVLTNMAFMESVHAKSYSSIFSTLCSTKQIDDAFDWSEQNPYL
QRKAQIIVDYRGDDALKRKASSVMLESFLFYSGFYLP MYWSSRGKLTNTADLIRLIIRDEAVHGYIIG
YKCQRGLADLTD AERADHREYTC ELLHTLYANEIDY AHDLYDELGWTD DVLPYMRYNANKALANLG
YQPAFDRDTCQVNP AVRAALDPGAGENHDF FSGSGSSYVMGTHQPTTDTDWDF

35 >Rv2092c helY helicase, Ski2 subfamily TB.seq 2349335:2352052 MW:99576 SEQ ID NO:213
VTELAELDRFTAELPFSLDDFQQRACSALERGHGVLCAPTGAGKTVVGEFAVHLALAAGSKCFYTT
PLKALSNQKHTDLTARYGRDQIGLLTGDLSVNGNAPVVMTEVLRNMLYADSPALQGLSYVMDE

VHFLADRMRGPVWEEVILQLPDDVRVWSLSATVSNAAEEFGGWIQTVRGDTTVVWDEHRPVPLWQHV
LVGKRMFDLFDYRIGEAGQPQVNRELLRHIAHRREADRMADWQPRRRGSGRPGFYRPPGRPEVI
AKLDAEGLLPITFVFSRAGCDAAVTQCLRSPLRLTSEEERARIAEVIDHRCGLADSDLAVLGYEYEW
REGLLRGLAAHHAGMLPAFRHTVEELFTAGLVKAVFATETLALGINMPARTVVLERLVKFNGEQHMP
5 LTPGEYTQLTGRAGRRGIDVEGHAVVIWHPEIEPSEVAGLASTRTFPLRSSFAPSYNMTINLVHRMGP
QQAHRLLSEQSFAQYQADRSVVGLVRGIERGNRILGEIAAELGGSDAPILEYARLRARVSELERAQARA
SRLQRRQAATDALAALRRGDIITITHGRRGGLAVVLESARDRDDPRPLVLTETHRWAGRISSADYSGTT
PVGSMTLPKRVEHRQPRVRRDLASALRSAAAGLVIPAARRVSEAGGFHDPELESSREQLRRHPVHT
SPGLEDQIRQAERYLRIERDNAQLERKVAATNSLARTFDRFVGLLTEREFIDGPATDPVVTDDGRLL
10 ARIYSESDLLVAECLRTGAWEGLKPAELAGVVSAVVYETRGGDGQGAPFGADVPTPRLRQALQTQS
RLSTTLRADEQAHRTPSREPDDGFVRVIYRWSRTGDLAAALAAADVNGSGSPLLAGDFVRWCRQV
LDLLDQVRNAAPNPELRATAKRAIGDIRRGVVAVDAG

>Rv2101 helZ helicase, Snf2/Rad54 family TB.seq 2360238:2363276 MW:111632

SEQ ID NO:214

MLVLHGFWSNSGGMRLWAEDSDLLVKSPSQALRSARHPFAAPADLIAGIHPGKPATAVLLLPSLR
APLDSPELIRLAPRPAARTDPMLLAWTVPVVDLDPTAALAAFDQPAPDVRYGASVDYLAELAVFAREL
VERGRVLPQLRRDTHGAAACWRPVLQGRDVVAMTSLVSAMPPVCRAEVGGHDPHELATSALDAMV
DAAVRAALSPMDLLPPRRGRSKRHRHAVEAWLTALTCPDGRFDAEPDELDALAEALRPWDDVGIGTV
20 GPARATFRLSEVETENEETPAGSLWRLEFLLQSTQDPSLLVPAEQAWNDDGSLRRWLD RPQELLLT
ELGRASRIFELVPALRTACPSGLELDADGAYRFLSGTAAVLDEAGFGVLLPSWWDRRRKGLGLVLSA
YTPVDGVVGKASKFGREQLVEFRWELAVGDDPLSEEEIAALTETKSPLIRLRGQWVALDTEQMRRGL
EFLERKPTGRKTTAEILALAASHPDDVDTPLEVTAVRADGWLGDLLAGAAAASLQPLDPPDGFTATLR
PYQQRGLAWLAFLSSGLGSLCLADDMGLGKTVQLLALETLESVQRHQDRGVGPTLLCPMSLVGN
25 WPQEAARFAPNLRVYAHHGARGLHGEALRDHLERTDLVVSTYTTATRDIDELAEYEWNRVVLDEAQ
AVKNSLSRAAKAVRRLRAAHRVALTGTPMENRLAELWSIMDFLNPGLLGSSERFRTRYAIPRIERHGH
EPAERLRASTRPYILRRLKTDPAIIDLPEKIEIKQYQCLTTEQASLYQAVVADMMEKIENTEGIERGN
VLAAMAKLKQVCNHPAQLLHRSPPVGRSGKVRLEEILEEILAEGDRVLCFTQFTEFAELLVPHLAAR
FGRAARDIAYLHGGTPRKRDEMVARFQSGDGPPIFLLSLKAGGTGLNLTAANHVVHLD RWWNPVAV
30 ENQATDRAFRIGQRRTVQVRKFICTGTLEEKIDEMIEEKKALADLVVTDGEGWLTSTRDLREVFAL
SEGAVGE

>Rv2110c prcB proteasome [beta]-type subunit 2 TB.seq 2369727:2370599 MW:30274

SEQ ID NO:215

VTWPLPDRLSINSLSGTPAVDLSSFTDFLRRQAPPELLPASISGGAPLAGGDAQLPHGTTIVALKYPGG
VVMAGDRRSTQGNMISGRDVRKYITDDYTATGIAGTAAVAVEFARLYAVELEHYEKLEGVPLTFAG
KINRLAIMVRGNLAAAMQGLLALPLLAGYDIHASDPQSAGRIVSFDAAGGWNIEEEGYQAVGSGSLFA

KSSMKKLYSQVTDGDSGLRVAVEALYDAADDDSATGGPDLVRGIFPTAVIIDADGAVDVPESRIAELA
RAIESRSGADTFGSDGGEK

>Rv2118c - = B2126_C1_165 (83.6%) TB.seq 2377471:2378310 MW:30091 SEQ ID NO:216

VSATGPFPSIGERVQLTDAKGRRYTMSLTPGAEFHTRGSIAHDAVIGLEQGSVVKSSNGALFLVLRPL
LV DYVMSMPRGPQVIYPKDAAQIVHEGDIFPGARVLEAGAGSGALTSLLRAVGPPAGQVISYEQRAD
HAEHARRNVSGCYGQPPDNWRLVVSDDLADSELPDGSVDRAVLDM LAPWEVLDAVSRLLVAGGVLM
VYVATVTQLSRIVEALRAKQCWTEPRAWETLQRGWNVVG LAVRPQHSMRGHTAFLVATRRLAPGA
VAPAPLGRKREGRDG

>Rv2144c - TB.seq 2404166:2404519 MW:12028 SEQ ID NO:217

MLIIALVLALIGLLALVFAVVTSNQLVAWVCIGASVLGVALLIVDALRERQQGGGADEADGAGETGVAEE
ADVDPYEEAPEESQAVDAGVIGSEEPSEEASEATEESAVSADRSDDSAK

>Rv2146c - TB.seq 2405667:2405954 MW:10805 SEQ ID NO:218

LVVFFQILGFALFIFWLLLIARVVVEFIRSFSDWRPTGVTTVILEIIMSITDPPVKVLRRLIPQLTIGAVRF
DLSIMVLLLVAFIGMQLAFGAAA

>Rv2147c - TB.seq 2406119:2406841 MW:27630 SEQ ID NO:219

VNSHCSHTFITDNRSRPRARRGHAMSTLHKVKAYFGMAPMEDYDDEYYDDRAPSRYARPRFDDDY
GRYDGRDYDDARSDSRGDLRGEPADYPPPGYRGGYADEPRFRPREFDRAEMTRPRFGSWLRNST
RGALAMDPRRMAMMFEDGHPLSKITTLRPKYSEARTIGERFRDGSFVIMDLVSM DNADAKRLVDF
AAGLAFALRG SFDKVATKVFLSPADVDVSPEERRRIAETGFYAYQ

>Rv2148c - TB.seq 2406841:2407614 MW:27694 SEQ ID NO:220

MAADLSAYPDRESELTHALAAMRSRLAAAAEAGRNVEIELLPITKFFPATDVAILFRLGCRSVGES
REQEASAKMAELNRLLAAAEELGHSGGVHWHMVGRIQRNKAGSLARWAHTAHSVDSSRLVTALDRA
VVAALAEHRRGERLRVYVQVSLDGDGSRGGVDSTTPGAVDRICAQVQESEGLELVGLMGIPPLDWD
PDEAFDRLQSEHNRVRAMFPHAIGLSAGMSNDLEVAVKHGSTCVRVGTALLGPRRLRSP

>Rv2150c ftsZ TB.seq 2408386:2409522 MW:38757 SEQ ID NO:221

MTPPHNYLAVIKVVGIGGGGVNAVNRMI EQGLKGVEFIAINTDAQALLMSDADVKLDVGRDSTRGLG
AGADPEVGRKAAEDAKDEIEELLRGADMVFTVAGEGGGTGTGGAPVVASIARKLGALT VGVVTRPF
SFEGKRRSNQAENGIAALRESCDTLIVIPNDRLLQM GDAAVSLMDAFRSADEVLLNGVQGITDLITP
GLINVDFAADVKGIMSGAGTALMGIGSARGEGRSLKAAEIAINSPLLEASMEGAQGVLM SIAGGSDLGL
FEINEAASLVQDA AHPDANIIFGTVIDDSLGD EVRVTVIAAGFDVSGPGRKPVMGETGGAHRIESAKA
GKLTSTLFEPVDAVS VPLHTNGATLSIGGDDDDVDVPPFMRR

>Rv2152c murC TB.seq 2410639:2412120 MW:51146 SEQ ID NO:222

VSTEQLPPDLRRVHVMVGIGGAGMSGIARILLDRGGLVSGSDAKESRGVHALRARGALIRIGHDASSL
DLLPGGATAVVTTHAAIPKTNPELVEARRRGIPVVLRAVLAKLMAGRTTLMVTGTHGKTTTTSMLIVA
5 LQHCGLDPSFAVGGELGEAGTNAHHGSGDCFVAEADESDGSLLQYTPHVAVITNIESDHLDFYGSVE
AYVAVFDSFVERIVPGGALVVCTDDPGGAALAQRATELGIRVLRYSVPGETMAATLVSWQQQGVG
AVAHIRLASELATAQGPRVMRLSVPGRHMALNALGALLAAVQIGAPADEVLDGLAGFEGVRRRFELV
GTCGVGKASVRVFDYAHHPTEISATLAAARMVLEQGDGGRCMVVFQPHLYSRTKAFAAEFGRALN
AADEVFVLDVYGAREQPLAGVSGASVAEHVTPMRYVPDFSAVAQQVAAAASPGDVIVTMGAGDVT
10 LLGPEILTALRVANRSAPGRPGVLG

>Rv2153c murG TB.seq 2412120:2413349 MW:41829 SEQ ID NO:223

VKDTVSQPAGGRGATAPRPADAASPCGSSPSADSVSVLAGGGTAGHVEPAMAVADALVALDPR
VRITALGTLRGLETRLVPQRGYHLELITAVPMPRKPGDLARLPSRVWRVREARDVLDVADVWV
15 GFGGYVALPAYLAARGLPLPPRRRRRIPVIHEANARAGLANRVGAHTADRVLSAVPDSGLRRAEVV
GVPVRASIAALDRAVLRAEARAHFGFPDDARVLLVFGGSQGAVSLNRAVSGAAADLAAAGVCVLHA
HGPQNVLELRRRAQGDPPYVAVPYLDRMELAYAAADLVICRAGAMTVAEVSAGVLPPIVPLPIGNG
EQRLNALPVVNAGGGMVADAALTPELVARQVAGLLTDPARLAAMTAAARVGHDAAGQVARAAL
AVATGAGARTTT

20

>Rv2154c ftsW TB.seq 2413349:2414920 MW:56306 SEQ ID NO:224

VLTRLLRRGTSDDTGSQTRGAEPVEGQRTGPPEASNPGSARPRTRFGAWLGRPMTSFHLIIAVAALL
TTLGLIMVLSASAVRSYDDDGSAWVIFGKQVLWTLVGLIGGYVCLRMSVRFMRRIAFSGFAITIVMLVL
VLVPGIGKEANGSRGWVWAGFSMQPSELAKMAFAIWGAHLLAARRMERASLREMLIPLVPAAVVAL
25 ALIVAQPDLGQTVSMGIILLGLLWYAGLPLRVFLSSLAAVVSAAILAVSAGYRSDRVRSWLNPENDP
QDSGYQARQAKFALAQQGIFGDGLGQGVAKWNYLPNAHNDFIFAIIGEELGLVGALGLLGLFGLFAY
TGMRIASRSADPFLRLLTATTTLWLWGQAFINIGYVIGLLPVTGLQLPLISAGGTSTAATLSLIGIANAAR
HEPEAVAALRAGRDDKVNRLRLPLPEPYLPPRLEAFDRDRKRANPQPAQTQPARKTPRTAPGQPAR
QMGLPPRPGSPRTADPPVRRSVHHGAGQRYAGQRRTRRVRALEGQRYG

30

>Rv2155c murD TB.seq 2414935:2416392 MW:49314 SEQ ID NO:225

VLDPLPGAPVLVAGGRVTGQAVAAVLTRFGATPTVCDPVMRLRPHAERGLPTVSSSDAVQQITG
YALVVASPGFSPATPLLAAGGVPIWGDVELAWRLDAAGCYGPPRSWLVTGTNGKTTTTSMLH
AMLIAGGRRRAVLGCGNIGSAVLDVLDEPAELLAVELSSQLHWAPSLRPEAGAVLNIAEDHLDWHATM
35 AEYTAAKARVLTGGVAVAGLDDSRAAALLDGSPAQVRVGFRLGEPARELGVRDAHLVDRAFSDDL
TLLPVASIPVPGPVGLDALAAALARSVGVPAGAIADAVTSFRVGRHRAEVVAVADGITYVDDSKAT
NPHAARASVLAYPRVWVIAGGLLKASLHAEVAAMASRLVGAVLIGRDRAAVAEALSRHAPDVPVVG

VVAGEDTGMPATVEVPVACVLDVAKDDKAGETVGAAVMTAAVAAAARRMAQPGDTVLLAPAGASFD
QFTGYADRGEAFATAVRAVIR

>Rv2156c murX TB.seq 2416397:2417473 MW:37714 SEQ ID NO:226

5 MRQILIAVAVAVTVSILLTPVLIRLFTKQGFHQIREDGPPSHHTKRGTPSMGGVAILAGIWAGYLGAH
LAGLAFDGEIGASGLLVGLLATALGGVGFIDDLIKIRRSRNLGLNKTAKTVGQITSAVLFGVLVLQFRN
AAGLTPGSADLSYVREIATVTLAPVLFVLCVVIVSAWSNAVNFTDGLDGLAAGTMAMVTAAYVLITF
WQYRNACVTAPGLGCYNVRDPLDLALIAAATAGACIGFLWWNAAPAKIFMGDTGSLALGGVIAGLSV
TSRTEILAVVLGALFVAEITSVVLQILTFRTTGRRMFRMAPFHHHFELVGWAETTVIIRFWLLTAITCGL
10 GVALFYGEWLAAGA

>Rv2157c murF TB.seq 2417473:2419002 MW:51634 SEQ ID NO:227

MIELTVAQIAEIVGGAVADISPQDAAHRRVTGTVEFDSRAIGPGGLFLALPGARADGHDHAASAVAAG
AAVVLAAARPVGVAIPVPPVAAPNVLAGVLEHDNDGSGAAVLAALAKLATAVAAQLVAGGLTIIGITGS
15 SGKTSTKDLMAAVLAPLGEVAPPGSFNNELGHPWTVLRATRRTDYILEMAARHHGNIAALAEIAPP
SIGVVLNVGTAHLGEFGSREVIAQTKAELPQAVPHSGAVVLNADDPAVAAMAKLTAARVVRVSRDNT
GDWWAGPVSLDELARPRFTLHAHDAQAEVRLGVCGDHQVTNALCAAVALECGASVEQVAAALTA
PPVSRHRMQVTTRGDGVTVIDDAYNANPDSMRAGLQALAWIAHQPEATRRSWAVLGEMAELEGDAI
AEHDRIGRLAVRLDVSRLVVVGTGRSISAMHHGAVLEGAWGSGEATADHGADRTAVNVADGDAALA
20 LLRAELRPGDVVLVKASNAAGLGAVADALVADDTCGSVRP

>Rv2158c murE TB.seq 2419002:2420606 MW:55310 SEQ ID NO:228

VSSLARGISRRRTEVATQVEAAPTGLRPNNAVGVRLAALADQVGAALAEQPAQRAVTEDRVTGVTIL
RAQDVSPGDLFAALTGSTTHGARHVGDALARGAVAVLTDPAQVAEIAAGRAAVPVLVHPAPRGVLGGL
25 AATVYGHPSERLTVIGITGTSGKTTTTYLVEAGLRAAGRVAGLIGTIGIRVGGADLPSALTTPEAPTLQA
MLAAMVERGVDTVMEVSSHALLGRVDGTRFAVGAFNTLSRDHLDHFHPSMADYFEAKASLFDPS
ALRARTAVVCIDDDAGRAMAARAADAITVSAADRPAAHWRATDVAPTDAGGQQFTAIDPAGVGHHIGI
RLPGRYNVANCLVALAILDTVGVSPEQAVPGLREIRVPGRLEQIDRGQGFLALVDYAHKPEALRSVLT
TLAHPDRRLAWVFGAGGDRDPGKRAPMGRIAAQLADLVVTDNPRDEDPTAIRREILAGAAEVGGD
30 AQVVEIADRRDAIRHAWARPGDWLIAGKGHETGQRGGGRVVPFDDRVELAAALEALERRA

>Rv2159c - TB.seq 2420632:2421663 MW:36377 SEQ ID NO:229

MKFVNHIIEPVAPRRAGGAVAEVYAEARREFGRLPEPLAMLSPDEGLLTAGWATLRETLVGQVPRG
RKEAVAAAASLRCPWCDAHTTMLYAAGQTDAAAILAGTAPAAGDPNAPYVAWAAGTGTPAGP
35 PAPFGPDVAAEYLGTAQVQFHFIAIRLVLLDETFLPGGPRAQQLMRRAGGLVFARKVRAEHRPGRST
RRLEPRTLPPDLAWATPSEPIATAFAALSHHLDTAPHLPPPTRQVVRVVGSWHGEPMPMSSRWNT

EHTAELPADLHAPTRLALLTGLAPHQVTDDDVAAARSLDDAALVGALAWAAFTAARRIGTWIGAAA
EGQVSRQNPTG

>Rv2163c pbpB TB.seq 2425049:2427085 MW:72506 SEQ ID NO:230

5 VSRAAPRRASQSQSTRPARGLRPPGAQEVGQRKRPGKTQKARQAQEQATKSRPATRSDVAPAGR
STRARRTRQVVDVGTGASVFRHRTGNAVILVLMVAATQLFFLQVSHAAGLRAQAAGQLKVTDV
QPAARGSIVDRNNDRLAFTIEARALTFQPKRIRRLQEEARKKTSAAPDPQQRLRDIAQEVAGKLNKPG
DAAAVLKKLQSDETFVYLARAVDPAVASAICAKYPEVGAERQDLRQYPGGSAAANVVGIDWDGHG
LLGLEDSLDAVLAGTDGSVTYDRGSDGVVPGSYRNRHKAVHGSTVVLTDNDIQFYVQQQVQQAK
10 NLSGAHNVSAVVLDAKTGEVLAMANDNTFDPSQDIGRQGDQKQLGNPAVSSPFEPGSVNKIVAASAVI
EHGLSSPDEVLPVPGSIQMGGVTVHDAWEHGVMPYTTTGVFGKSSNVGTLMLSQRVGPERYDML
RKFGGLGQRTGVGLPGESAGLVPPIDQWSGSTFANLPIGQGLSMTLLQMTGMYQAIANDGVRVPPRII
KATVAPDGSRTTEPRPDDIRVVSQAQTAQTVRQMLRAVVQRDPMGYQQGTGPTAGVPGYQMAGKT
GTAQQINPGCGCYFDDVYWITFAGIATADNPRYVIGIMLDNPARNSDGAPGHSAAPLFHNIAGWLMQ
15 RENVPLSPDPGPPLVLQAT

>Rv2165c - TB.seq 2428236:2429423 MW:42498 SEQ ID NO:231

VQTRAPWSLPEATLAYFPNARFVSSDRDLGAGAAPGIAASRSTACQTWGGITVADPGSGPTGFGHV
PVLAQRCFELLTPALTRYYPDGSQAVLLDATIGAGGHAERFLEGLPGLRLIGLDRDPTALDVARSLV
20 RFADRLTLVHTRYDCLGAALAESGYAAVGSVDGILFDLGVSSMQLDRAERGFAATDAPLDMRMDP
TTPLTAADIVNTYDEAALADILRRYGEERFARRIAAGIVRRRAKTPFTSTAEVLVALLYQAIPAPARRVGG
HPAKRTFQALRIAVNDELESLRTAVPAALDALAIGGRIAVLAYQSLEDRIKRVFAEAVASATPAGLPV
ELPGHEPRFRSLTHGAERASVAEIERNPRSTPVRLRALQRVEHRAQSQQWATEKGDS

25 >Rv2166c - TB.seq 2429428:2429856 MW:15912 SEQ ID NO:232

MFLGTYTPKLDDKGRLLTPAKFRDALAGGLMVTKSQDHS LAVYPRAAFEQLARRASKAPRSNPEAR
AFLRNLAAGTDEQHPDSQGRITLSADHRRYASLSKDCVIGAVDYLEIWDAQAWQNYQQIHEENFSA
ASDEALGDIF

>Rv2197c - TB.seq 2461505:2462146 MW:22481 SEQ ID NO:233

30 MVSRYSA YRRGPDVISPDVIDRILVGACAAVWL VFTGVSVA AVALMDLGRGFHEMAGNPHTTWL
YAVIVVSALVIVGAIPVLLRARRMAEAEPATRPTGASVRGGRSIGSGHPAKRAVAESAPVQHADA FEV
AAEWSSEAVDRIWL RGTVVL TSAIGIALIAVAAATYLM AVGHDGPSWISYGLAGVV TAGMPVIEWLYA
RQLRRVVAPQSS

>Rv2198c - TB.seq 2462149:2463045 MW:30955 SEQ ID NO:234

35 MSGPNPPGREPDEPESEPVSDTGDERASGNHLPPVAGGGDKLPDQTGETDAYSRAVSAPESHEV
TGGPYVPADLRLYDYDDYEESDLDDELAAPRPWWVVGVAIIAVALVSVSLLVTRPHTSKLATG
DTTSSAPPVQDEITTTKPAPPPPPPPPPPTTEIPTATETQTVTVTPPPPPPPATTTAPPPATTTTAAAP

PPTTTTPTGPRQVTYSVTGTKAPGDIISVTYVDAAGRRTQHNVYIPWSMTVTPISQSDVGSVEASSL
FRVSKLNCSITTS DGT V LSSNSNDGPQTSC

>Rv2199c - TB.seq 2463234:2463650 MW:14866 SEQ ID NO:235

5 MHIEARLFEFVAFFVWTAVLYGVLTSMFATGGVEWAGTTALALTGGMALIVATFFRFVARRLDSRPE
DYEGAEISDGAGELGFFSPHSWWPIMVALSGSVAAGVIALWLPWLIAAGVAFILASAAGLVFEYYVGP
EKH

>Rv2200c ctaC TB.seq 2463661:2464749 MW:40449 SEQ ID NO:236

10 VTPRGPGRLLQRLSQCRPQRGSGGPARGRLQLALAAMLGALAVTVSGCSWSEALGIGWPEGITPEA
HLNRELWIGAVIASLAVGVVWGLIFWSAVFHRKKNTDTLPRQFGYNMPELVLTVIPFLIISVLFYFT
VVVQEKMQLQAKDPEVIDITSFQWNWKFQYQVRNFKDGTLYDGADPERKRAMVSKPEGKDKYGE
ELVGPVRGLNTEDRTYLNFDKVELTGTSTEIPVLVLP SGKRIEFQMASADVIHAFWWPEFLFKRDVMP
NPVANNSVNVFQIEEITKTGAFVGHCAEMCGTYHSMNNFEVRVTPNDFKAYLQQRIDGKTNAEALR
15 AINQPPLAVTTHPFDTRRGELAPQPVG

>Rv2427c proA g-glutamyl phosphate reductase TB.seq 2724231:2725475 MW:43746
SEQ ID NO:237

20 MTVPAPSQDLRQEVHDAARRARVAARRLASLPTTVKDRALHAADELLAHRDQILAANAEDLNAAR
EADTPAAMLDRLSLNPQRVDGIAAGLRQVAGLRDPVGEVLRGYTLPNGLQLRQQRVPLGVVGMIE
GRPNVTVD AFG LTLKSGNAALLRGSSSAKSNEALVAVLRTALVGLELPADAVQLLSAADRATVTHLI
QARGLVDVIPRGAGLIEAVVRDAQVPTIETGVGNCHVYVHQAADLDVAERILLNSKTRRPSVCNA
AETLLVDAAIAETALPRLLAALQHAGVTVHLDPDEADLRREYLSLDIAVAVVDGVDAAIAHINEYGTGH
TEAIVTTNLDAQRFTEQIDAAVMVNASTAFTDGEQFGFGAEIGISTQKLHARGPMGLPELTSTKWI
25 AWGAGHTRPA

>Rv2438c - similar to YHN4_YEAST P38795 TB.seq 2734793:2737006 MW:80492
SEQ ID NO:238

30 MGLLGGQSGPRVSGSPVGSIPVNAICQQRGGFHGVERGYSAGDSGVLTS LGDNERTMNFYSA
YQHGFVRVAACTHHTTIGDPAANAASVLDMARACHDDGAALAVFPELTLSGYSIEDVLLQDSLLDAV
EDALLDLVTESADLLPVLVVGAPLRHRHRIYNTAVVIHRGAVLGVPKSYLPTYREFYERRQMAPGD
GERGTIRIGGADVAFGTDLLFAASDLPGFVLHVEICEDMFVPMPPSAEALAGATVLANLSGSPITIGR
AEDRRLLARSASARCLAAYVYAAAGEGESTTDLAWDGQTMIWENGALLAESERFPKGVRRSVADVD
TELLRSERLRMGTFFDDNRRHHRELTESFRRIDFALDPPAGDIGLLREVERFPFVPADPQRLQQDCYE
35 AYNIQVSGLEQRLRALDYPKVVIGVSGGLDSTHALIVATHAMDREGRPRSDILAFALPGFATGEHTKN
NAIKLARALGVTSEIDIGDTARLMLHTIGHYPYSVGEKVYDVTFENVQAGLRDYLFRIANQRGGIVLG
TGDLSELALGWSTYGVGDQMSHYNVNAGVPKTLIQHLIRWWISAGEFGGEKVGEVLQSVLDTEITPELI

PTGEEELQSSEAKVGPFALQDFSLFQVLRYGFRPSKIAFLAWHAWNDAERGNWPPGFPKSERPSYS
LAEIRHWLQIFVQRFYSFSQFKRSALPNGPKVSHGGALSPRGDWRAPSDMSARIWLDQIDREVPKG

>Rv2439c proB glutamate 5-kinase TB.seq 2737118:2738245 MW:38789 SEQ ID NO:239

5 MRSPHRDAIRTARGLVWKVGTTALTTPSGMFDAGRLAGLAEAVERRMKAGSDVIVSSGAIAAGIEPL
GLSRRPKDLATKQAAASVGQVALVNSWSAAFARYGRTVGQVLLTAHDISMVQHTNAQRTLDRRLA
LHAVAIVNENDTVATNEIRFGDNDRLSALVAHLVGADALVLLSDIDGLYDCDPRKTADATFIPEVSGPA
DLDGVVAGRSSHLLGTGGMASKVAAALLAADAGVPVLLAPAADAATALADASVGTVFAARPARLSAR
RFWVRYAAEATGALTDAGAVRAVVRQRRSLAAGITAVSGRFCGGDVVELRAPDAAMVARGVVAY
10 DASELATMVGRSTSELPGELRRPVHADDLVAVSAKQAKQV

>Rv2440c obg Obg GTP-binding protein TB.seq 2738248:2739684 MW:50430
SEQ ID NO:240

VPRFVDRVIHTRAGSGGNGCASVHREKFKPLGGPDGGNGGRGGSIVFVDPQVHTLLDFHFRPHL
15 TAASGKHGMGNNRDGAAGADLEVKVPEGTVVLDENGRLLADLVGAGTRFEAAAGGRGGLGNAALA
SRVRKAPGFALLGEKGQSRDLTLELKTADVGLVGFPSSAGKSSLVSAISAAKPKIADYPFTTLVPNLG
VVSAGEHAFTVADVPLIPGASRGRGLGLDFLRHIERCAVLVHVDCATAEPGRDPISDIDALETELA
CYPTLQGDAAALGDLAARPRAVLNKIDVPEARELAEFVRDDIAQRGWPVFCVSTATRENQLPLIFGL
SQMISDYNAARPVAVPRRPVIRPIVDDSGFTVEPDGHGGFVSGARPERWIDQTNFDNDEAVGYL
20 ADRLARLGVEEELLRLGARSGCAVTIGEMTFDWEPTPAGEPVAMSGRGTDPRLDSNKRVGAAER
KAARSRRREHGDG

>Rv2441c rpmA 50S ribosomal protein L27 TB.seq 2739773:2740030 MW:8969
SEQ ID NO:241

25 MAHKKGASSSRNGRDSAAQRLGVKRYGGQVVKAGEILVRQRGTFKHPGVNVGRGGDDTLFAKTAG
AVEFGIKRGRKTVSIVGSTTA

>Rv2442c rplU 50S ribosomal protein L21 TB.seq 2740048:2740359 MW:11152
SEQ ID NO:242

30 MMATYAIVKTGGKQYKVAVGDVVKVEKLESEQGEKVSLPVALVVDGATVTTDAKALAKVAVTGEVLG
HTKGPKIRIHKFNKTGYHQRQGHRRQQLTVLKVTGIA

>Rv2448c valS valyl-tRNA synthase TB.seq 2747596:2750223 MW:97822 SEQ ID NO:243

MLPKSWDPAAMESAIYQKWLDAGYFTADPTSTKPAYSIVLPPPNVTGSLHMGHALEHTMMDALTRR
35 KRMQGYEVLWQPGTDHAGIATQSVVEQQLAVDGGTKEDLGRELFDKVDWKRESGGAIGGQMR
RLGDGVDWSRDRFTMDEGLSRAVRTIFKRLYDAGLIYRAERLVNWSPLQTAISDLEVNYRDVEGEL
VSFRYGSLLDSQPHIVATTTRVETMLGDTAIAVHPDDERYRHLVGTSLAHPFVDRELAIVADEHVDPE

FGTGAVKVT PAHDPNDFEIGVRHQLP MPSILDTKGRIVDTGTRFDGMDRFEARVAVRQALAAQGRV
VEEKRPYLHSVGHSESRSGEPIEPRLSLQWWWVRVESLAKAAGDAVRNGDTVIHPASMEPRWFSWVD
DMHDWCISRQLWWGHRIPWYGP DGEQVCVGPDET PPQGW EQDPDVLDTWFSSALWPFSTL GW
PDKTAELEKFYPTSVLVTGYDILFFWVARM MMFGTFVGDDAAITLDGRRGPQVPFTD VFLHGLIRDE
5 SGRKMSKSKGNVIDPLDWVEMFGADALRFTLARGASPGGDLAVSEDAVRASRNFGTKLFNATRYAL
LNGAAPAPLPSPNELTDADR WILGRLEEVR AEVDSAFDGYEFSRACESLYHFAWDEFCDWYLELAK
TQLAQGLTHTTAVLAAGLDTLLRLLHPVIPFLTEALWLALTGRESLVSADWPEPSGISVDLVAAQRIND
MQKLVT E VRRFRSDQGLADRQKVPARMHGVRDSDLSNQVAAVTSLAWLTEPGPDFEPSVSLEVR L
GPEMNRTVVVELDTSGTIDVAAERRRLEKELAGAQKELASTAAKLANADFLAKAPDAVIAKIRDRQRV
10 AQQETERITTRLAALQ

>Rv2482c plsB2 TB.seq 2786915:2789281 MW:88284 SEQ ID NO:244

VTKPAADASAVLTAEDTLVLASTATPVEMELIMGWL GQQRARHPDSKFDILKLPPRNAPPAALTALVE
QLEPGFASSPQSGEDRSIVPVRVIWLPPADRSRAGKVAALLPGRDPYHPSQRQRRILRTDPRRAR
15 VVAGESAKVSEL RQQWRDTTVAEHKR DFAQFVSRRALLALARA EYRILGPQYKSPRLVKPEMLASA
RFRAGLD RIPGATVEDAGKMLDELSTGWSQVSVDLVSVLGR LASRGFDPEFDYDEYQVAAMRAALE
AHPAVLLFSHRSYIDGVVVPVAMQDNRLPPVHMF GGINSFGLMGPLMRRSGMIFIRRNIGNDPLYK
YVLKEYVG YVVEKRFNLSWSIEGTRSR TGKMLPPKLG LMSYVADAYLDGRSDDILLQGV SICFDQLH
EITEYAAYARGAEKTPEGLRWLYNFIKAQGERNF GKIYVRFPEAVSMRQYLGAPHGELTQDPAAKRL
20 ALQKMSFEVAVRILQATPVTATGLVSALLTTRGTALT DQLHHTLQDSL DYLERKQSPVSTSALRLR
SREGVRAAADALSNGHPVTRVDSGREPVWYIAPDDEHAAAFYRNSVIHAFLET SIVELALAHAKHAE
GDRVAAFWAQAMRLRDLLKFD FYFADSTAFRANIAQEMAWHQDWEDHLGVGGNEIDAMLYAKRPL
MSDAMLRVFFEAYEIVADVL RDAPPD IGPEELTELALGLGRQFVAQGRVRSSEPVSTLLFATARQVAV
DQELIAPAADLAERRVAFRRELRNLRD FDYVEQIARNQFVACEFKARQGRDRI
25

>Rv2509 - putative oxidoreductase TB.seq 2824676:2825479 MW:28014 SEQ ID NO:245

MPIPAPSPDARAVVTGASQ NIGAALATELAARGHHLIVTARREDVLT ELAARLADKYRVTVDVRPADL
ADPQERSKLAD ELAARPISILCANAGTATFGPIASLDLAGEKTQVQLNAVAVHDLTLAVLP GMIERKAG
GILISGSAAGNSPIPYNATYAATKAFVNTFSES LRGELRGSGVHVTVLAPGPV RTELPDASEASLVEKL
30 VPDFLWISTEHTARVSLNALERNKMRVVPGLTSKAMSVASQYAPRAIVAPIVGAFYKRLGGS

>Rv2524c fas fatty acid synthase TB.seq 2840124:2849330 MW:326226 SEQ ID NO:246

VTIHEHDRVSADRGGDSPHTTHALVDRLMAGEPYAVAFGGQGS AWLETLEELVSATGIETELATLVG
EAELLLDPVTDELIVVRPIGF EPLQWVRALAAEDPVPSDKHLTSAAVSVPGVLLTQIAATRALARQGM
35 DLVATPPVAMAGHSQGV LAVEALKAGGARDVELFALAQLIGAAGTLVARRRGISVLGDRPPMVSVTN
ADPERIGRLLDEFAQDVRTLPPVLSIRNGRRRAVITGTPEQLSRFELYCRQISEKEEADRKNKVRGG
DVFSPVFEPVQVEVGFHTPRLSDGIDIVAGWAEKAGLDVALARELADAILRKVDWVDEITRVHAAGA

RWILDLGPGDILTRLTAPVIRGLGIGIVPAATRGGQRNLFTVGATPEVARAWSSYAPTVVRLPDGRVK
LSTKFTRLTGRSPILLAGMTPTTVDAKIVAAAANAGHWAELAGGGQVTEEIFGNRIEQMAGLLEPGRT
YQFNALFLDPYLWKLQVGGKRLVQKARQSGAIDGVVISAGIPDLDEAVELIDELGDIGISHVFKPGT
IEQIRSVIRIATEVPTKPVIMHVEGGRAGGHHSWEDLDDLLATYSELRSRANITVCVGGGIGTPRRAA
5 EYLSGRWAQAYGFPLMPIDGILVGTAAATKESTTSPSVKRMLVDTQGTDDQWISAGKAQGGMASSR
SQLGADIHEIDNSASRCGRLLDEVAGDAEAVAERRDEIIAAMAKTAKPYFGDVADMTYLQWLRRYVE
LAIGEENSTADTASVGSPLADTWRDRFEQMLQRAEARLHPQDFGPIQTLFTDAGLLDNPQQAIAAL
LARYPDAETVQLHPADVPFFVTLCCKTLGKPVNFVPVIDQDVRWWRSDSLWQAHDARYDADAVCIIP
GTASVAGITRMDEPVGELLDRFEQAAIDEVLGAGVEPKDVASRRLGRADVAGPLAVLDAPDVRWA
10 GRTVTNPVHRIADPAEWQVHDGPENPRATHSSSTGARLQTHGDDVALSVPVSGTWVDIRFTLPANTV
DGGTPVIATEDATSAMRTVLAIAAGVDSPEFLPAVANGTATLTVDWHPERVADHTGVTATFGEPLAP
SLTNVPDALVGPCWPAVFAAIGSAVTDGTGEPVVEGLLSLVHLDHAARVVGQLPTVPAQLTVTATAAN
ATDMDMGRVVPVSVVVTGADGAVIATLEERFAILGRTGSAELADPARAGGAVSANATDTPRRRRRDV
TITAPVDMRPFVAVSGDHNPIHTDRAAALLAGLESPIVHGMWLSAAAQHAVTATDGQARPPARLVG
15 WTARFLGMVRPGDEVDFRVERVGDQGAIEVDVAARVGSDDLMSASARLAAPKTVYAFPGQGIQHK
GMGMEVRARSKAARKVWDATDKFTRDTLGFSVLHVVRDNPTSIASGVHYHHPDGVLYLTQFTQVA
MATVAAAQVAEMREQGAFVEGAIAACGHSVGEYALACVTGIYQLEALLEMFHRSKMHMDIVPRDEL
GRSNYRLAARPSQIDLDDADVPFVAGIAESTGEFLEIVNFNLRSQYAIAGTVRGLEALEAEVERRR
ELTGGRRSFILVPGIDVPFHSRVLRVGVAEFRRSLDRVMPRDADPDLIIGRYIPNLVPRLFTLDRDFIQ
20 EIRDLVPAEPLDEILADYDTWLRERPREMARTVFIELLAWQFASPVRWIETQDLLFIEEAAGGLGVERF
VEIGVKSSPTVAGLATNTLKLPEYAHSTVEVLNAERDAAVLFATDTPPEPEPEEDEPVAESPAPDVS
EAAPVAPAASSAGPRPDDLVDAAATLALIALSAKMRIDQIEELDSIESITDGASSRRNQLLVDLGSE
LNLGAIDGAAESDLAGLSQVTKLARTYKPYGPVLSDAINDQLRTVLGPSGKRPGAIAERVKKTWELG
EGWAKHVTVEVALGTREGSSVRGGAMGHLHEGALADAASVDKVIDAASVAARQGVSVLPSAG
25 SGGGATIDAAALSEFTDQITGREGVLASAARVLVGLGLDDPVNALPAAPDSELIDLVTaelGADWPR
LVAPVFDPKKAVVFDWRASAREDLVKLWLDEGDIDADWPRLAERFEGAGHVATQATWWQGKS
LAAGRQIHASLYGRIAAGAENPEPGRYGGEVAVVTGASKGSIAASVVARLLDGGATVIATTSKLDEER
LAFYRTLYRDHARYGAALWLVAANMASYSDDVALVEWIGTEQTESLGPQSIHIKDAQPTLLFPFAAP
RVVGDLEAGSRAEMEMKVLLWAVQRLIGGLSTIGAERDIASRLHVLPGPSNRMFGGDGAYGEA
30 KSALDAVVSRWHAESSWAARVSLAHLIGWTRGTGLMGHNDIVAAVEEAGVTTYSTDEMAALLD
LCDAESKVAARSPIKADLTGGLAEANLDMELAAKAREQMSAAAVIDEAEAPGAIAALPSPPRGF
TPAPPPQWDDLVDPADLVVIVGGAEIGPYGSSRTRFEMEVENELSAAGVLELAWTTGLIRWEDDP
QPGWYDTESGEMVDESELVQRYHDAVVQRVGIREFVDDGAIDPDHASPLLVSFLEKDFAFVVSSE
ADARAFVEFDPEHTVIRPVPDSTDWQVIRKAGTEIRVPRKTKLSRVVGGQIPTGFDPTVWGISADMA
35 GSIDRLAVWNMVATVDAFLSSGFSPAEMRYVHPSLVANTQGTGMGGGTSMTMYHGNLLGRNKP
NDIFQEVLPNIAAHVQSYVGSYGAMIHPVAACATAAVSVEEGVDKIRLGKAQLVAGGLDDLTLEGII
GFGDMAATADTSMCGRGHDSKFSRPNDRRLGFVEAQGGGTILLARGDLALRMGLPVLAVVAFA

QSFQDGVHTSIPAPGLGALGAGRGGKDSPLARALAKLGVAADDVAVISKHDTSTLANDPNETELHER
LADALGRSEGAPLFVVSQKSLTGHAKGGAAVFQMMGLCQILRDGVIPPNRSLDCVDELAGSAHFV
WVRDTLRLGGKFPLKAGMLTSLGFGHVSGLVALVHPQAFIASLDPADRADYQRRADARLLAGQRRRL
ASAIAGGAPMYQRPQDGRFRDHHAPERPEASMLLNPAARLGDGEAYIG

5

>Rv2555c alaS alanyl-tRNA synthase TB.seq 2873772:2876483 MW:97326 SEQ ID NO:247
VQTHEIRKRFLDHFVKAGHTEVPSASVILDDPNLLFVNAGMVQFVPFFLGQRTPPYPTATSIQKCI RTP
DIDEVGITTRHNTFFQMAGNFSFGDYFKRGAIELAWALLTNSLAAGGYGLDPERIWTTVYFDDDEAV
RLWQEVAGLPAERIQRGMADNYWSMGIPGPCGPSSEIYYDRGPEFGPAGGPVISED RYLEVWNL
10 VFMQNERGEGTTKEDYQILGPLPRKNIDTGMGVERIALVLQDVHNVYETDLLRPVIDTVARVAARAYD
VGNHEDDVRYRIIADHSRTAAILIGDGVSPGNDGRGYVLRRLRRVIRSAKLLGIDAAVGDLMATVRN
AMGPSYPELVADFERISRIAVAEETA FNRTLASGSRLFEEVASSTKKSGATVLSGSDAFTLHDTYGFP I
ELTLEMAAETGLQVDEIGFRELMAEQRRRAKADAAARKHAHADLSAYRELVDAGATEFTGFDELRS
QARILGIFVDGKRVPVVAHGVAGGAGEGQ RVELVLDRTPLYAESGGQIADEGTISGTGSSEAARAAV
15 TDVQKIAKTLWVHRVNVESGEFVEGDTVIAAVDPGWRRGATQGHSGTHMVHAALRQVLGPNAVQA
GSLNRPGYLRFDFNWQGPLTDDQRTQVEEVTNEAVQADFEVRTFTEQLDKAKAMGAIALFGESYPD
EVRVWEMGGPFSLELCGGTHVSNTAQIGPV TILGESSIGSGVRRVEAYVGLDSFRHLAKERALMAGL
ASSLKVPSEEVPARVANLVERLRAAEKELERVRMASARAAATNAAAGAQ RIGNVRLVAQRMSGGMT
AADLRSLIGDIRGKLGSEPAVVALIAEGESQTPYAVAANPAAQDLGIRANDLVKQLAVAVEGRGGGK
20 ADLAQSGSKNPTGIDAALDAVRSEIAVIARVG

20

>Rv2580c hisS histidyl-tRNA synthase TB.seq 2904822:2906090 MW:45118 SEQ ID NO:248
VTEFSSFSAPKGVDPDYPPDSAQFVAVRDGLLAAARQAGYSHIELPIFEDTALFARGVGESTDVVSKE
MYTFADRGDRSVTLRPEGTAGVVRVIEHGLDRGALPVKLCYAGPFFRYERPQAGRYRQLQQVGV
25 EAIGVDDPALDAEVIAIADAGFRSLGLDGRLEITSLGDESCRPQYRELLQEFLEGLDLEDDRTRRAGI
NPLRVLDDKRPEL RAMTASAPVLLDHLSDVAKQHFDTVLAHLDALGVPPYINPRMVRGLDYYTKTAF
EFVHDGLGAQSGIGGGGGRYDGLMHQLGGQDL SGIGFGLGVDRTVLALRAEGKTAGDSARCDVFGV
PLGEAAKLRLAVLAGRLRAAGVRVDLAYGDRGLKGAMRAAARSGARVALVAGDRDIEAGTVAVKDL
TTGEQVSVSMDSVVAEVISRLAG

30

>Rv2614c thrS threonyl-tRNA synthase TB.seq 2941190:2943265 MW:77123 SEQ ID NO:249
MSAPAQAPAGVDGGDPSQARIRVPAGTTAATAVGEAGLPRRGTPDAIVVVRDADGNLRDLSWVPD
VDTDITPVAANTDDGRSVIRHSTAHLVAQAVQELFPQAKLGIGPPITDGFYYDFDVPEPFTPEDLAAL E
KRMQRIVKEGQLFDRRVYESTEQARAELANEPYKLELVDDKSGDAEIMEVGGDEL TAYDNLNPRTR
35 ERVWGDLCRGP HIPTTKHIPAFKLTRSSAAYWRGDQKNASLQRIYGTAWESQEALDRHLEFIEEAQR
RDHRKLGVELDLFSFPDEIGSGLAVFHPKGGIVRRELEDYSRRKHTEAGYQFVNSPHITKAQLFHTSG
HLDWYADGMFPPMHIDAEYNADGSLRKPGQDYLLKPMNCPMHCLIFRARGRSYRELPLRLFEGTV

35

YRYEKSGVHGLTRVRGLTMDDAHIFCTRDQMRDELRSLLRFVLDLLADYGLTDFYLELSTKDPEKF
VGAEVWEEATTVLAEVGAESGLELVPDPGGAIFYGPKISVQVKDALGRTWQMSTIQLDFNFPERF
GLEHTAADGTRHRPVMHRLFGSIEFFGILTEHYAGAFPAWLAPVQVVGIPVADEHVAYLEEVATQ
LKSHGVRAEVDASDDRMMAKKIVHHTNHKVPFVLAGDRDVAAGAVSFRFGDRTQINGVARDDAVAA
5 IAWIADRENAVPTAELVKVAGRE

>Rv2697c dut deoxyuridine triphosphatase TB.seq 3013683:3014144 MW:15772 SEQ ID NO:250
VSTTLAIVRLDPGLPLPSRAHDGDAGVDLYSAEDVELAPGRRALVRTGVAVAVPFGMVGLVHPRSGL
ATRVGLSIVNSPGTIDAGYRGEIKVALINLDPAAPIVVHRGDRIAQLLVQRVELVELVEVSSFDEAGLAS
10 TSRGDGGHGGSSGGHASL

>Rv2782c pepR protease/peptidase, M16 family (insulinase) TB.seq 3089045:3090358 MW:47074
SEQ ID NO:251
MPRRSPADPAAALAPRRTTLPGGRLRVTEFLPAVHSASVGVWVGVSDEGATVAGAAHFLEHLLF
15 KSTPTRSAVDIAQAMDAVGGELNAFTAKEHTCYAHVLGSDLPLAVDLVADVNLNGRCAADDVEVER
DVLLEEIAMRDDDPEADALADMFLAALFGDHPVGRPVIGSAQSVSVMTRAQLQSFHLRRYTPERMVV
AAAGNVHDHGLVALVREHFGSRLVRGRRPVAPRKGTRVNGSPRLTLVSRDAEQTHVSLGIRTPGR
GWEHRWALSVLHTALGGGLSSRLFQEVRETRGLAYSVYSALDLFADSGALSVYAACLPERFADVMR
VTADVLESVARDGITEAECGIAKGSRLGGVLGLDSSSRMSRLGRSELNYGKHRSIEHTLRQIEQVT
20 VEEVNAVARHLLSRRYGA AVLGP HGSKRSLPQQLRAMVG

>Rv2783c gpsi pppGpp synthase and polyribonucleotide phosphorylase TB.seq
3090339:3092594 MW:79736 SEQ ID NO:252
MSAAEIDEGVFETTATIDNGSFGTRTIRFETGRLALQAAGAVVAYLDDDNMLLSATTASKNPKEHFDF
FPLTVDVEERMYAAGRIPGSFFRREGRPSTDAILTCRLIDRPLRPSFVDGLRNEIQIVVTILSLDPGDLY
25 DVLAINAASASTQLGGLPFSGPIGGVRVALIDGTWVGFPPTVDQIERAVFDMVAVGRIVEGDVAIMMVE
AEATENVVELVEGGAQAPTESVVAAGLEAAKPFIAALCTAQQELADAAGKSGKPTVDFPVFPDYGED
VYYSVSSVATDELAALTTIGGKAERDQRIDEIKTQVVQRLADTYEGREKEVGAALRALTKKLVRQRILT
DHFRIDGRGITDIRALSAEVAVVPRAHGSALFERGETQILGVTTLDMIKMAQQIDSLGPETSKRYMHH
YNFPPFSTGETGRVGSPPKRREIGHGALAERALVPVLPSVEEFYAIRQVSEALGSNGSTSMGSVCAS
30 TLALLNAGVPLKAPVAGIAMGLVSDDIQVEGAVDGVVERRFVTLTDILGAEDAFGDMDFKVAGTKDFV
TALQLDTKLDGIPSQVLGALEQAKDARLTILEVMAEAIDRPDEMSPYAPRVTTIKVPVDKIGEVIGPK
GKVINAITEETGAQISIEDDGTVFVGATDGPSAQAIDKINAIANPQLPTVGERFLGTVVKTDFGAFVS
LLPGRDGLVHISKLGKKGKRIAKVEDVNVGDKLRVEIADIDKRGKISLILVADEDSTAAATDAATVTS

>Rv2793c truB tRNA pseudouridine 55 synthase TB.seq 3102364:3103257 MW:31821
35 SEQ ID NO:253
MSATGPGIIVIDKPAGMTSHDVVGRCRRIFATRRVGHAGTLDPMATGVLVIGIERATKILGLLTAAPKS
YAATIRLGQTTSTEDAEGQVLQSVPAKHLTIEAIDAAMERLRGEIRQVPSSVSAIKVGGRRAYRLARQ

GRSVQLEARPIRIDRFELLAARRRDQLIDIDVEIDCSSGTIYRALARDLGDALGVGGHVTALRRTRVGR
FELDQARSLDDLAERPALSLSLDEACLLMFARRDLTAAEASAAANGRSPLAVGIDGVYAACDADGRVI
ALLRDEGSRTSRVAVLRPATMHPG

5 >Rv2797c - TB.seq 3105619:3107304 MW:58761 SEQ ID NO:254

VPLTVADIDRWNAQAVREVFAASARA EVTFEASRQLAALSIFANS GGKTA EAAAHHNAGIRRD LDA
HGNEALAVARAADRAADGIVKVQSELAALRHAAAAAELTIDALINRVVPIPLRSTEAQWARTLAKQT
ELQAE L DAIMAEANAVDEELASAVNMADGDAIPADSGPPVGPEGLTPTQLASDANEERLREERARL
QAHLE RLQAEYDQLSVRAARDYHNGILDGDAVGRLAALTDELSAARGRLGELDAVDEALS RAPETYL
10 TQLQIPEDPNQQVLA AVAVGNPD TAANVS VTPGVGSTTRGALPGMVTEARDLRSEVIRQLNAAGK
PASVATI AWMGYHPPPNPLDTGSAGDLWQTM TDGQA HAGAADLSRYLQQVRANNPSGHLTVLGHS
YGSLTASLALQDLDAQSAHPVNDVVFY GSPGLELYSPAQLGLDHGHAYVMQAPHDLITNLVAPLAPL
HGWGLDPYLT PGFTELSSQAGFDPGGIWRDGVYAHGDYPRSF L DAAGQPQLRMSGYNLAAIAAGL
PDNTVGPPLLPPILGGGMPAAPGPALRGGR

15

>Rv2864c ponA2 TB.seq 3175454:3177262 MW:63015 SEQ ID NO:255

MVTKTTLASATSGLLLLAVVAMSGCTPRPQGPGPA AEKFFAALAIGDTASAAQLSDNPNEAREALNA
AWAGLQAAHLDAQVLSAKYAEDTGT VAYRFSWHL PKDRIW TYDGQLKMARDEGRWHVRWTTSG L
HPKLGEHQTFALRADPPRRASVNEVG GTDVLVPGYLYHYS LDAGQAGRELFGTAHAVVGALHPFDD
20 TLNDPQLLAEQASSSTQPLDLVTLHADD SNRVAAAIGQLPGVITPQAELLPTDKHFAPAVLNDVKKA
VDEL D GKAGWRVSVNQNGVDVSVLHEVAPSPASSVSITLDRVVQNA AQHAVNTRGGKAMIVVIK
PSTGEILAI AQNAGADADGPVATTGLYPPGSTFKMITAGAAVERDLATPETLLGCPGEIDIGHRTIPNY
GGFDLGVPMSRAFASSCNTTFAELSSRLPPRGLTQAARRYGIGLDYQVDGITTVTG SVPPTVDLAE
RTEDGFGQGKVLASPFGMALVAATVAAGKTPVPQLIAGRPTAVEGDATPISQKMIDALRPMMLVVT
25 NGTAKEIAGCGEVFGKTGEAEFPGGSHSWFAGYRGDLAFASLIVGGGSSEYAVRMTKVMFESLPPG
YLA

>Rv2868c gcpE TB.seq 3179368:3180528 MW:40451 SEQ ID NO:256

VTVGLGMPQPPAPT LAPRRATRQLMVG NVGVGSDHPVSVQSMCTTKTHDVNSTLQQIAELTAAGC
DIVRVACPRQEDADALAEIARHSQIPVADIHFQPRYIFAAIDAGCAAVRVNPGNIKEFDGRVGEVAKA
30 AGAAGIPIRIGVNAGSLDKRFMEKYGKATPEALVESALWEASLFEEHGF GDIKISVKHNDPVMVAAY
ELLAARCDYPLHLGVTEAGPAFQGTIKSAVAFGALLSRGIGDTIRVSLSAPPVEEVKVG NQVLESNL
RPRSLEIVSCPSCGRAQVDVYTLANEV TAGLDGLDVPLRVAVMGC VVNGPGEAREADLGVASGN GK
GQIFVRGEVIKTVPEAQIVETLIEEAMRLAAEMGEQDPGATPSGSPIVTVS

>Rv2869c - TB.seq 3180548:3181759 MW:42835 SEQ ID NO:257

MMFVTGIVLFALAILISVALHECGHMWVARRTGMKVRRYFVGFGPTLWSTRRGETEYGVKAVPLGG
35 FCDIAGMTPVEELDPDERDRAMYKQATWKRVAVLFAGPGMNLAI CLVLIYAIALVWGLPNLHPPTRAV
IGETGCVAQEVSQGKLEQCTGPGPAALAGIRSGDVVVKVGDTPVSSFDEMAAAVRKSHGSPIVVE

RDGTAIVTYVDIESTQRWIPNGQGGELQPATVGAIGVGAARVGPVRYGVFSAMPATFAVTGDLTVEV
GKALAALPTKVGALVRAIGGGQRDPQTPISVVGASIIIGD TVDHGLWVAFWFFLAQLNLILAAINLLPL
LPFDGGHIAVAVFERIRNMVRSARGKVAAAPVNYLKLLPATYVVLVLVVG YMLLTVTADLVNPIRLFQ
>Rv2870c - TB.seq 3181770:3183077 MW:45324 SEQ ID NO:258

5 VATGGRVWIRRRGDNEVVAHNDEVNSTDGRADGRLRVVVLGSTGSIGTQALQVIADNPDRFEVVG
LAAGGAHLDTLLRQRAQTGVTNIAVADEHAAQRVGDIPYHGSDAATRLVEQTEADVNLALVGALGL
RPTLAALKTGARLALANKESLVAGGSLVLRARPQGIVPVDSEHSALAQCLRGGTPDEVAKLVLTA
GGPFRGWSAADLEHVTPEQAGAHPTWSMGPMNTLNSASLVNKGLEVIETHLLFGIPYDRIDVVVHP
QSIHSMVTFIDGSTIAQASPPDMKLPISLALGWPRRVSGAAAACDFHTASSWEFEPLD TDVFPAVEL
10 ARQAGVAGGCMTAVYNAANEEAAAAFLAGRIGFPAIVGIIADVLHAADQWAVEPATVDDVLDAQRWA
RERAQRAVSGMASVAIASTAKPGAAGRHAHLERS

>Rv2922c smc member of Smc1/Cut3/Cut14 family TB.seq 3234189:3238055 MW:139610
SEQ ID NO:259

15 VGAGSRFPLVDPLPSVGARPDRLRGQPRRRTRAGGRPGSARCVPEAAAAAGRHD TGPRRQSRR
RLVAVDGADHRVQRAVIWPLVYLKSLTLKGFSFAAPTTLRFEPGITAVVGPNGSGKSNVVDALAWV
MGEQGAKT LRGGKMEDVIFAGTSSRAPLGRAEVTVSIDNSDNALPIEYTEVSITRRMFRDGASEYEIN
GSSCRLMDVQELLSDSGIGREMHVIVGQKLEELQSRPEDRRAFIEEAAGVLKHKRKEKALRKLD T
MAANLARLTDLTTELRRQLKPLGRQAEAAQRAAAIQADLRDARLRLAADDLVSRRAEREAVFQAEAA
MRREHDEAAARLAVASEELAHAESAVAE LSTRAESIQTWFGLSALAERVDATVRIASERAHHLDIEP
20 VAVSDTDPRKPEELEAEAQVAVAEQQLLAELDAARARLDAARAELADRERRAAEADRAHLAAVRE
EADRREGLARLAGQVETMRARVESIDESVARL SERIEDAAMRAQQTRA EFETVQGRIGELDQGEVG
LDEHHERTVAALRLADERVAELQSAERAERQVASLRARIDALAVGLQRKDGAAWLAHNRSGAGLF
GSIAQLVKVRSGYEALAAALGPAADALVDGLTAAGSAVSALKQADGGRAVLVLSDWPAPQAPQS
ASGEMLP SGAQWALDLVESPPQLVGAMIAMLSGVAVVNDLTEAMGLVEIRPELRAVTV DGDVLVAG
25 WVSGGSDRKLSTLEVTSEIDKARSELAAAEALAAQLNAALAGALTEQSARQDAAEQALALNESDTAI
SAMYEQLGRLGQEARAAEEEWNRLLQQRTEQEAVRTQTLD DVIQLETQLRKAQETQRVQVAQPIDR
QAISAAADRARGVEVEARLAVRTAEERANAVRGRADSLRRAAAEREARVRAQQARAARLHAAAVA
AAVADCGRLLAGRLHRAVDGASQLRDASAAQRQRLAAMA AVRDEVNTLSARVGELTDSLHRDEL
ANAQAALRIEQLEQMVLEQFGMAPADLITEYGP HVALPPTLEMAEFEQARERGEQVIAPAPMPFDR
30 VTQERRAKRAERALAELGRVNPLALEEFAALEERYNFLSTQLEDVKAARKDLLGVVADVDARILQVFN
DAFVDVEREFRGVFTALFPGGEGRLRLTEPDDMLTTGIEVEARPPGKKITRLSLLSGGEKALTAVAML
VAIFRARPSPFYIMDEVEAALDDVNLRRLLSLFEQLREQSQIIITHQKPTMEVADALYGV TMQNDGITA
VISQRMRGQQVDQLVTNSS

>Rv2925c mc RNase III TB.seq 3239829:3240548 MW:25400 SEQ ID NO:260

35 MIRS RQPLLDALGV DLPDELLSLALTHRSYAYENGGLPTNERLEFLGDAVLGLTITDALFH RHDPDRSE
GDLAKLRASVNTQALADVARRLCAEGLGVHVL LGRGEANTGGADKSSILADGMESLLGAIYLQHGM

EKAREVILRLFGPLLDAAPTLGAGLDWKTSLQELTAARGLGAPSYLVTSTGPDHDKEFTAVVVVMDSEYGSVGRSKKEAEQKAAAAAWKALEVLDNAMPKGTSA

>Rv2934 ppsD TB.seq 3262245:3267725 MW:193317 SEQ ID NO:261

5 MTSLAERAAQLSPNARAALARELVLAGTTFTDICEPVAVVGIGCRFPGNVTPGESFWQLLADGVDT
IEQVPPDRWDADAFYDPDPSASGRMTTKWGGFVSDVDAFDADFFGITPREAVAMPQHRMLLEVA
WEALEHAGIPDLSGTRTGVMGLSSWDYTIVNIERRADIDAYLSTGTPHCAAVGRIAYLLGLRGPA
VAVDTACSSSLVAIHLACQSLRLRETDVALAGGVQLTSPFTAIALSKWSALSPTGRCNSFDANADGF
VRGEGCGVWLKRLADAVRDQDRVLAVVRGSATNSDGRSNGMTAPNALAQRDVITSALKLADVTPD
10 SVNYVETHGTGTVLGDPIEFESLAATYGLKGQGESPCALGSKVTNIGHLEAAAGVAGFIKAVLAVQR
GHIPRNLHFTRWNPADASATRLFPPTESAPWPAAGPRRAAVSSFGLSGTNAHVVEQAPDTAVAA
AGGMPYVSALNVSGKTAARVASAAVLADWMSGPGAAAPLADVAHTLNRHRARHAKFATVIARDRA
EAIAGLRALAAGQPRVGVVDCDQHAGGPGRVFVYSGQGSQWASMGQQLLANEPAFAKAVAEIDPI
FVDQVGFSLQQTLDGDEVGIDRIQPVLVGMQLALTELWRSYGVIPDAVIGHSMGEVSAAVWAGALT
15 PEQGLRVITTRSRLMARLSGQGAMALLELDADAAEALIAGYQVTLAVHASPRQTVIAGPPEQVDTVI
AAVATQNRLARRVEVDVASHHPIDPILPELRSALADLTPQPPSIPIISTTYESAQPVADADYWSANLRN
PVRFHQAVTAAGVDHNTFIEISHPVLTHALTDLDPDGSHTVMSTMNRELDQTLFYHAQLAAVGVA
ASEHTTGRLVDLPPTPWHHQRFWTDRSAMSELAATHPLLGAHIEMPRNGDHWQTDVGTEVCPW
LADHKVFGQPIMPAAGFAEIALAAASEALGTAADAVAPNIVINQFEVEQMLPLDGHTPLTTQLIRGGDS
20 QIRVEIYSRTRGGEFCRHATAKVEQSPRECAHAHPEAQGPATGTTVSPADFYALLRQTGQHHGPAF
AALSRIURLADGSAETEISIPDEAPRHPGYRLHPVVLDAALQSVGAAIPDGEIAGSAEASLYPVSFETIR
VYRDIGRHVRCRAHLTNLDGGTGKMGRIVLINDAGHIAAEVDGIYLRVVERRAVPLPLEQKIFDAEWT
ESPIAAVPAPEPAAETTRGSWLVLADATVDAPGKAQAKSMADDFVQQWRSPMRRVHTADIHDESAV
LAFAETAGDPEHPPGVVVFVGGAASSRLDDELAARDTVWSITTVRAVVGTVHGRSPRLWLVTG
25 GGLSVADDEPGTPAAASLKGLVRVLAFEHPDMRTTLVDLITQDPLTALSALRNAGSGSRHDDVIA
WRGERRFVERLSRATIDVSKGHPVVRQGASYVVTGGLGGLGLVVARWLVDRGAGRVLGGRSDPT
DEQCNVLAELQTRAEIVVVRGDVASPGVAEKLIETARQSGGQLRGVHAAVIEDSLVFSMSRDNLE
RVWAPKATGALRMHEATADCELDWWLGFSSAASLLGSPGQAAYACASAWLDALVGWRRASGLPA
AVINWGPWSEVGVAQALVGSVLDTISVAEGIEALDSSLAADRITGVARLRADRALVAFPEIRSISYFT
30 QVVEELDSAGDLGDWGGPDALADLDPGEARRAVTERMCARIAAVMGYTDQSTVEPAVPLDKPLTEL
GLDSLMAVRIRNGARADFGVEPPVALILQGASLHDLTADLMRQLGLNDPDPALNNADTIRDRARQRA
AARHGAAMRRRPKPEVQGG

>Rv2946c pks1 TB.seq 3291503:3296350 MW:166642 SEQ ID NO:262

35 VISARSAEALTAQAGRLMAHVQANPGLDPIDVGCSLASRSVFEHRAVVGASREQLIAGLAGLAAGE
PGAGVAVGQPGSVGKTVVFPQGGAQRIGMGRELYGELPVFAQAFDAVADELDRHLRLPLRDVIW
GADADLLDSTFAQPALFAVEVASFAVLRDWGVLPDFVMGHSVGEIAAAHAAGVLTADAAMLVVA

RGRLMQALPAGGAMVAVAASEDEVEPLLGEGVGIAAINAPESVVISGAQAAANAIADRFAAQGRRVH
QLAVSHAFHSPLMEPMLEEFARVAARVQAREPQLGLVSNVTGELAGPDFGSAQYWVDHVRPVRF
ADSARHLQTLGATHFIEAGPGSGLTGSIEQSLAPAEAMVVSMLGKDRPELASALGAAGQVFTTGVPV
QWSAVFAGSGGRRVQLPTYAFQRRRFWETPGADGPADAAGLGLGATEHALLGAVVERPDSDEVVL
5 TGRSLADQPWLADHVNGVWLFPGAGFVELVIRAGDEVGCALIEELVLAAPLVMHPGVGVQVQVWV
GAADESGHRAVSVYSRGDQSQGWLLNAEGMLGVAAAETPMDLSVWPPEGAESVDISDGYAQLAE
RGYAYGPAFQGLVAIWRRGSELF AEVVAPGEAGVAVDRMGMPAVLDAVLHALGLAVEKTQASTET
RLPFCWRGVSLHAGGAGRVRARFASAGADAISVDVCDATGLPVLTVRSLVTRPITAEQLRAAVTAAG
GASDQGPLEVWSPISVVS GGANGSAPPAPVSWADFCAGSDGDASVVVWELESAGGQASSVGS
10 VYAATHTALEVLQSWLGADRAATLVVLTHGGVGLAGEDISDLAAA VWMARSAQAENPGRIVLIDT
DAAVDASVLAGVGEPQLLVRRGGTVHAPRLSPAPALLALPAESA WRLAAGGGGTLEDLVIQPCPEV
QAPLQAGQVRVAVAAVGVNFRDVVAALGMYPGQAPPLGAEGAGVLETGPEVTDLAVGDAVMGFL
GGAGPLAVVDQQLVTRVPQGW SFAQAAAVPVVFLTAWYGLADLAEIKAGESVLIHAGTGGVGMAAV
QLARQWGVVEFVTASRGKWDTLRAMGFDDDHIGDSRTCEFEKFLAVTEGRGVDVVLDSLAGEFV
15 DASLRLLVRRGGRFLEMGKTDIRDAQEIAANYPGVQYRAFDLSEAGPARMQEMLAEVRELFDTRELH
RLPVTTWDVRCAPAAFRFMSQARHIGKVLTMP SALADRLADGT VITGATGAVGGVLARHLVGAY
GVRHLVLASRRGDRAEGAAELAADLTEAGAKVQVACDVADRAAVAGLFAQLSREYPPVRGVIHAA
GVLDDAVITSLTPDRIDTVLRAK/DAAWN LHQATSDLDLSMFALCSSIAATV GSPGQGNYSANAFLD
GLAAHRQAAGLAGISLAWGLWEQPGGMTAHLSSRDLARMSRSLAPMSPAEAVELFDAALAI DHPL
20 AVATLLDRAALDARAQAGALPALFSGLARRRRRQIDDTGDATSSKSALAQRLHGLAADEQLELLVG
LVCLQAAAVLGRPSAEDVDPDTEFGDLGFDLSL TAVELRNRLKTATGLTLPPTVIFDHPTPTAVAEYVA
QQMSGSRPTESGDPTSQVVEPAAAEVSVHA

>Rv3014c ligA DNA ligase TB.seq 3372545:3374617 MW:75258 SEQ ID NO:263

VSSPDADQTAPEVL RQWQALAEVREHQFRYYVRD APIISDAEFDELLRRLEALEEQHP ELRTPDSP
25 TQLVGGAGFATDFEPVDHLERMLSLDNAFTADELA AWAGRIHAEVGDAAHYLCELKIDGVALSLVYR
EGRLTRASTRGDGRGTGEDVT LNARTIADVPERLTPGDDYPVPEVLEVRGEVFFRLDDFQALNASLVE
EGKAPFANPRNSAAGSLRQKDP AVTARRRLRMICHGLGHVEGFRPATLHQAYLALRAWGLPVSEHT
TLATDLAGVRERIDYWGEHRHEVDHEIDGVVVKVDEVALQRR LGSTS RAPRWAIAYKYPPEEAQTKL
LDIRVNVGRTGRITPFAFMT PVKVAGSTVGQATLHNASEIKRKGVLIGDTVIRKAGDVIPEVLGPVVE
30 LRDGSEREFIMPTTCPEGCSPLAPEKEGDADIRCPNARGCPGQLRERV FHVASRNGLDIEVLGYEAG
VALLQAKVIADEGELFALTERDLLRTDLFR TKAGELSANGKRLLVNLDKAKAAPLWRVLVALSIRHVGP
TAARALATEFGSLDAIAAASTDQLAAVEGVGPTIAAAVTEWFAVDWHREIVDKWRAAGVRMVDERD
ESVPRTLAGLTIVVTGSLTGFSRDDAKEAIVARGGKAAGSVSKKTNYV VAGDSPGSKYDKAVELGVPI
LDEDGFRLLADGPASRT

35 >Rv3025c - Nifs-like protein TB.seq 3383885:3385063 MW:40948 SEQ ID NO:264

MAYLDHAATTPMHAAIEAMAAVQRTIGNASSLHTSGRSARRRIEEARELIADKLGARPSEVIFTAGG
TESDNLAVKGIYWARRDAEPHRRRIVTTEVEHHAVLDSVNWLV EHEGAHVTWLPTAADG SVSATAL

REALQSHDDVALVSVMWANNEVGTLPIAEMSVMAMEFGVPMHSDAIQAVGQLPLDFGASGLSAMS
VAGHKFGGPPGVGALLLRDVTVCVPLMHGGGQERDIRSGTPDVASAVGMATAAQIAVDGLEENSAR
LRLLRDRLVEGVLAIEDDVCLNGADDPMLAGNAHFTFRGCEGDALLMLLDANGIECSTGSACTAGV
AQPSHVLIAMGVDAASARGSLRLSLGHTSVEADVDAALEVLPGAVARARRAALAAAGASR

5

>Rv3080c pknK serine-threonine protein kinase TB.seq 3442656:3445985 MW:119420

SEQ ID NO:265

MTDVDPHATRRDLVPNIPAEELLEAGFDNVEEIGRGGFGVVYRCVQPSLDRAVAVKVLSTDLDNRNLE
RFLREQRAMGRLSGHPHIVTLQVGVLAGGRPFIWMPYHAKNSLETIRRHGPLDWRETLSIGVKLA
10 GALEAAHRVGTLHRDVKPGNILLTDYGEPLTDFGIARIAGGFETATGVIAGSPAFTAPEVLEGASPTP
ASDVYSLGATLFCALTGHAAYERRSGERVIAQFLRITSQPIPDRLKQGLPADVAAAIERAMARHPADR
PATAADVGEELRDVQRRNGVSVDEMPLPVELGVERRRSPEAHAAHRHTGGGTPTVPTPTPATKY
RPSVPTGSLVTRSRLTDILRAGGRRRLILIHAPSGFGKSTLAAQWREELSRDGAAVAWLTIDNDDNNE
VWFLSHLLESIRRVPTLAESLGHVLEEHGDDAGRYVLTSLIDEIHENDDRIAVVIDDWHRVSDSRTQ
15 AALGFLLDNGCHHLQLIVTSWSRAGLPVGRRLRIGDELAIDSALRFDTDEAAALLNDAGGLRLPRAD
VQALTTSTDGWAAALRLAALSRLGGGDATQLLRGLSGASDVIHEFLSENVLDLEPELREFLLVASVT
ERTCGGLASALAGITNGRAMLEEAHGRFLQRTEDDPNWFRLFHQMFAFLHRRLERGGSHRVAEL
HRRASAWFAENGYLHEAVDHALAAGDPAVAVDLVEQDETNLPEQSKMTTLLAIVQKLPTSMVVSRA
RLQLAIWANILLQRPAPATGALNRFETALGRAELPEATQADLRAEADVLRAVAEVFADRVERVDLL
20 AEAMSRPDTLPPRPVPGTAGNTAALAAICRFEFAEVYPLLDWAAPYQEMMGPFGTVYAQCLRGMAAR
NRLDIVAALQNFRATAFEVGTAVGAHSHAARLAGSLLAELLYETGDLGAGRLMDESILLGSEGGAVD
YLAARYVIGARVKAQGDHEGAADRLSTGGDTAVQLGLPRLAARINNERIRLGIALPAVAADLLAPR
TIPRDNGIATMTAELDEDSAVRLLSAGDSADRDQACQRAGALAAIDGTRRPLAALQAQILHIETLAAT
GRES DARNELAPVATKCAELGLSRLLDVAGLA

25

>Rv3106 fprA adrenodoxin and NADPH ferredoxin reductase TB.seq 3474004:3475371

MW:49342 SEQ ID NO:266

MRPYYIAIVGSGPSAFFAAASLLKAADTTEDLDMAVDMLEMLPTPWGLVRSGVAPDHPKIKSISKQFE
KTAEDPRFRFFGNVVVGEHVQPGELSERYDAVIYAVGAQSDRMLNIPGEDLPGSIAAVDFVGWYNA
HPHFQVSPDLSGARAVVIGNGNVALDVARILLTDPDVLARTDIADHALESRLPRGIQEVVIVGRRGPL
30 QAAFTTLELRELADLDGVDVIDPAELDGITDEDAAVGVCKQNIKVLRGYADREPRPGHRRMVFR
FLTSPIEIKGKRKVERIVLGRNELVSDGSGRVAAKDTGEREELPAQLVVRVSVGYRGVPTPGLPFDDQ
SGTIPNVGGRINGSPNEYVVGWIKRGPTGVIGTNKKDAQDTVDTLIKNLGNAKEGAECKSFPEDHAD
QVADWLAARQPKLVTSAHWQVIDAFERAAGEPHGRPRVKLASLAELLRIGLG

35

>Rv3235 - TB.seq 3611296:3611934 MW:22659 SEQ ID NO:267

MMASNQTAAQHSSATLQQAPRSIDDAGGCPLTISPIANSPGDTFAVTPVVEYEPNIPPCGQSSH
AARRPHTPQLARRQPIRPSGRAPAAVTSTAKSPRLRQAGTFADAALRRVLEVIDRRRPVQGLRPLLA

PGLVDSVLAVSRTAAGHQGAAMLRRIRLTPAGPDTADTAAEVFGTYSRGDRIHAIACRVEQRPAGN
ETRWLMVALHIG

>Rv3255c manA mannose-6-phosphate isomerase TB.seq 3635040:3636263 MW:43340

SEQ ID NO:268

5 VELLRGALRTYAWGSRTAIAEFTGRPVPAAHPEAELWFGAHPGDPAWLQTPHGQTSLEALVADPE
GQLGSASRARFGDVLPLVKVLADEPLSLQAHPSAEQAVEGYLREERMGIPVSSPVRNYRDTSHK
PELLVALQPFEALAGFREARTTELLRALAVSDLPFIDLLSEGSDADGLRALFTTWITAPQPDIDVLV
PAVLGAIQYVSSGATEFGAEAKTVLELGERYPGDAGVLAALLNRISLAPGEAIFLPAGNLHAYVRG
FGVEVMANSNDNVLRGGLTPKHVDVPELLRVLDFAPTPKARLRPPIRREGLGLVFETPTDEFAATLLVL
10 DGDHLGHEVDASSGHDGPQILLCTEGSATVHGKCGSLTLQRGTAAWVAADDGPIRLTAGQPAKLFR
ATVGL

>Rv3264c rmlA2 glucose-1-phosphate thymidyltransferase TB.seq 3644897:3645973 MW:37840

SEQ ID NO:269

LATHQVDVAVLVGGKGTRLRPLTLSAPKPMLPTAGLPFLTHLLSRIAAAGIEHVILGTSYKPAVFEEAF
15 GDGSALGLQIEYVTEEHPLGTGGGIANVAGKL RNDTAMVFNGDVL SGADLAQLLDFHRSNRADVTL
QLVRVGDPRAFGCVPTDEEDRVVAFLEKTEDPPTDQINAGCYVFERNVIDRIPQGREVSVEREVFPA
LLADGDCKIYGYVDASYWRDMGTPEDFVRGSADLV RGIAPSPALRGHRGEQLVHDGA AVSPGALLI
GGTVVGRGAEIGPGTRLDGAVIFDGVRVEAGCVIERSIIGFGARIGPRALIRDGVIGDGADIGARCELL
SGARVWPVGVLDPDGGIRYSSDV

20

>Rv3368c - TB.seq 3780334:3780975 MW:23734 SEQ ID NO:270

MTLNLSVDEVLT TTRS VRKRLDFDKPVPRDVLMECLELALQAPTGSNSQGWQWVFVEDAAKKKAIA
DYYLANARGYLSGPAPEYPDGDTRGERMGRVRDSATYLAEHMHRAPVLLIPCLKGREDES AVGGVS
FWASLFP AVWSFCLALRSRGLGSCWTTLHLLDNGEHKVADVLGIPYDEYSQGGLLPIAYTQGIDFRP
25 AKRLPAESVTHWNGW

>Rv3382c lytB1 TB.seq 3796447:3797433 MW:34667 SEQ ID NO:271

MAEVFVGPAQGYASGEVTLLASPRSF CAGVERAIETVKRVLDAEGPVYVRKQIVHNTVVVAELR
DRGAVFVEDLDEIPDPPPPGAVVVFSAHGVSPAVRAGADERGLQVVDATCPLVAKVHAEARFAAR
30 GDTVVFIGHAGHEETEGTLGVAPRSTLLVQTPADVAALNLPEGTQLSYLTQTTLALDETADVIDALRA
RFPTLGQPPSEDICYATTNRQALQSMVGECDVVLVIGSCNSSNSRRLVELAQRS GTPAYLIDGPDDI
EPEWLSSVSTIGVTAGASAPPRLVGQVIDALRGYASITVVERS IATETVRFGLPKQVRAQ

>Rv3418c groES 10 kD chaperone TB.seq 3836985:3837284 MW:10773 SEQ ID NO:272

35 VAKVNIKPLEDKILVQANEAETTTASGLVIPDTAKEKPQEGTVAVGPGRWDEDEGEKRIPLDVAEGDT
VIYSKYGGTEIKYNGEEYLILSARDVLAVVSK

>Rv3423c alr TB.seq 3840193:3841416 MW:43357 SEQ ID NO:273

VKRFWENVGKPNDDTDGRGTTSLAMTPISQTPGLLAEAMVDLGAIEHNVRVLREHAGHAQLMAVVK
ADGYGHGATRVAQTALGAGAAELGVATVDEALALRADGITAPVLAWLHPPGIDFGPALLADVQVAVS
SLRQLDELLHAVRRTGRTATVTVKVDGLNRNGVGPAQFPAMLTALRQAMAEDAVRLRGLMSHMV
5 YADKPDDSINDVQAQRFTAFLAQAREQQVRF EVAHLSNSSATMARPDLTDFDLVRPGIAVYGLSPVPA
LGDMGLVPAMTVKCAVALVKSIRAGEGVSYGHTWIAPRDTNLALLPIGYADGVFRSLGGRLEVLINGR
RCPGVGRICMDQFMVDLGPGLDVAEGDEAILFGPGIRGEPTAQDWADLVGTIHYEVVTSRGRITR
TYREAENR

10 >Rv3490 otsA [alpha].-trehalose-phosphate synthase TB.seq 3908232:3909731 MW:55864
SEQ ID NO:274

MAPSGGQEAQICDSETFGDSDFVWANRLPVDLERLPDGSTTWKRSPGGLVTALEPVLRRRRGAW
VGWPGVNDDGAEPDLHVLGDGPIIQDELELHPVRLSTTDIAQYYEGFSNATLWPLYHDVIVKPLYHRE
WWDRYVDVNQRFAEAAASRAAAHGATVWVQDYLQVLVPKMLRMLRPDLTIGFFLHIPFPVELFMQ
15 MPWRTEIIQGLLGADLVGFHLPGGAQNFLILSRRLVGTDTSRGTVGVRSRFGAAVLGSRIRVGAFFI
SVDSGALDHAARDNRIRRRAREIRTELGNPRKILLGVDRLDYTKGIDVRLKAFSELLAEGRVKRDDTV
WQLATPSRERVESYQTLRNDIERQVGHINGEYGEVGHPPVHYLHRPAPRDELIAFFVASDVMLVTP
LRDGMNLVAKEYVACRSDLGGALVLSFTGAAAELRHAYLVNPHDLEGVKDGIEEALNQTEEAGRR
RMRSLLRRQVLAHDVDRWAQSFLDALAGAHPRGQG

20

>Rv3598c lysS lysyl-tRNA synthase TB.seq 4041423:4042937 MW:55678 SEQ ID NO:275

VSAADTAEDLPEQFRIRRDKRARLLAQGRDPYPVAVPRTHLAEVRAAHPDLPIDTATEDIVGVAGRV
IFARNSGKLCFATLQDGDGTQLQVMISLDKVGQAALDAWKADVDLGDIVYVHGAVISSRRGELSVLA
DCWRIAASLRPLPVAHKEMSEESRVRQRYVDLVRPEARAVARLRIAVVRAIRTAQRRGFLEVETP
25 VLQTLAGGAAARPFATHSNALDIDLRIAPELFLKRCIVGGFDKVFELNRVFRNEGADSTHSPEFSM
LETYQTYGTYDDSAVVTRELIQEVADEAIGTRQLPLPDGSVYDIDGEWATIQMYPSSLVALGEEITPQT
TVDRLRGIADSLGLEKDPAlHDNRGFGHGKLIELWERTVGKSL SAPTFVKDFPVQTTPLTRQHR SIP
GVTEKWDLYLRGIELATGYSELSDPVVQRERFADQARAAAAGDDEAMVLDEDFLAALEYGMPPCTG
TGMGIDRLLMSLTGLSIRETVLFPIVRPHSN

30

>Rv3600c - similar to Bacillus subtilis protein YacB TB.seq 4043041:4043856 MW:29274
SEQ ID NO:276

VLLAIDVRNTHTVVGLLSGMKEHAKVVQQWRIRTESEVTADELALTIDGLIGEDSERLTGTAALSTVPS
VLHEVRIMLDQYWPSVPHVLEPGVRTGIPLLVDNPKEVGADRVNCLAAYDRFRKAAIIVDFGSSICV
35 DVVSAKGEFLGGAIAPGVQVSSDAAAARS AALRRVELARPRSVVGKNTVECMQAGAVFGFAGLVDG
LVGRIREDVSGFSVDHDAIVATGHTAPLLLPELHTVDHYDQHLTLQGLRLVFERNLEVQRGRLLKTAR

>Rv3606c folK 7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase TB.seq

4048181:4048744 MW:20732 SEQ ID NO:277

MTRVVLVSGSNLGDRLARLSVADGLGDALIAASPIYEADPWGGVEQGQFLNAVLIADDPTCEPREW
LRRAQEFERAAGRVRGQRWGPRNLDVDLIACYQTSATEALVEVTARENHLLPHPLAHLRAFLIPW
5 IAVDPTAQLTVAGCPRPVTRLLAELEPADRDSVRLFRPSFDLNSRHPVSRAPES

>Rv3607c folX may be involved in folate biosynthesis TB.seq 4048744:4049142 MW:14553

MADRIELRGLTVHGRHGVYDHERVAGQRFVIDVTWIDLAEAAANSDDLADTYDYVRLASRAAEIVAG
PPRKLIETVGAEIADHVMDDQRVHAVEVAVHKPQAPIPQTFFDDVAVVIRRSRRGGRGWVVPAGGAV

10 >Rv3608c folP dihydropteroate synthase TB.seq 4049138:4049977 MW:28812 SEQ ID NO:278

VSPAPVQVMGVLNVTDDSFSDGGCYLDLDDAVKHGLAMAAAGAGIVDVGGESSRPGATRVDPAVE
TSRVIPVVKELAAQGITVSIDTMRADVARAALQNGAQMVNDVSGGRADPAMGPLLAADV PWVLMH
WRAVSADTPHPVRYGNVVAEVRADLLASVADAVAAGVDPARLVLDPLGLGFAKTAQHNWAILHALP
ELVATGIPVLVGASRKRLGALLAGPDGVMRPTDGRDGTATAVISALAALHGAWGVRVHDVRSVDAI
15 KVVEAWMGAERIERDG

>Rv3609c folE GTP cyclohydrolase I TB.seq 4049977:4050582 MW:22395 SEQ ID NO:279

MSQLDSRSASARIRVFDQQRAEAAVRELLYAIGEDPDRDGLVATPSRVARSYREMFAGLYTDPDSVL
NTMFDEDHDELVLVKEIPMYSTCEHHLVAFHGVAVHGYIPGDDGRVTGLSKIARLVLDLYAKRPQVQE
RLTSQIADALMKKLDPRGVIVVIEAEHLCEMAMRGVRKPGSVTTTTSAVRGLFKTNAASRAEALDLILRK

20 >Rv3610c ftsH inner membrane protein, chaperone TB.seq 4050601:4052880 MW:81987

MNRKNVTRTITAIWVLLGWSFFYFSDDRGYKPVDTSAITQINGDNVKSQAQIDDREQLRLILKKG
NNETDGSEKIVITKYPTGYAVDLFNALSAKNAKVSTVNNQGSILGELLVYVLP LLLLVLGLFVMFSRMQG
GARMGFGFGKSRKQLSKDMPKTTFADVAGVDEAVEELYEIKDFLQNPSTRYQALGAKIPKGVLLYGP
PGTGKTLARAVAGEAGVPFFTISGSDFFVEMFVGVGASRVRLDFEQAKQNSPCIIFVDEIDAVGRQR
25 GAGLGGGHDEREQTLNQLLVEMDGFGRAGVILIAATNRPDILDALLRPGRFDRQIPVSNPDLAGR
RAVLRVHSGKGKPMADADLDGLAKRTVGMTGADLANVINEAALLTARENGTVITGPALEEAVDRVIG
GPRRKGRRISEQEKKITAYHEGGHTLAAWAMPDIEPIYKVITILARGRTGGHAAVAVPEEDKGLRTRSEMI
AQLVFAMGGRAAEELVFREPTTGAVSDIEQATKIARSMVTEFGMSSKLGAVKYGSEHGDPFLGRMT
GTQPDYSHEVAREIDEEVRKLIEAAHTEAWEILTEYRDVLDLAGELEKETLHRPELESIFADVEKRP
30 RLTMFDDFGGRIPSDKPKPIKTPGELAIERGEPWPQVPPEPAFKAAIAQATQAAEAARSDAGQTGHGA
NGSPAGTHRSGDRQYGSTQPDYAGAPAGWHAPGWPPRSSHRPSYSGEPAPTYPGQPYPTGQADP
GSESSAEQDDEVSRTKPAHG

>Rv3671c - TB.seq 4112322:4113512 MW:40722 SEQ ID NO:280

MTPSQWLDIAVLAVAFIAAISGWRAGALGSMLSFGGVLLGATAGVLLAPHIVSQISAPRAKLFAALFLIL
35 ALVVVGEVAGVWLGRAVRGAIRNRPIRLIDSVIGVGVQLVVVLTAAWLLAMPLTQSKEQPELAAAVKG
SRVLARVNEAAPTWLKTVPKRLSALLNTSGLPAVLEPFSRTPVIPVASPDALVNNPVVAATEPSVVKI
RSLAPRCQKVLEGTGFVISPDVRMTNAHVAGSNNVTYAGDKPFEATVVSYPDPSVDVAILAVPHLP

PPPLVFAAEPAKTGADVWVLGYPGGGNFTATPARIREAIRLSGPDYGDPEPVTRDVYTIRADVEQGD
SGGPLIDLNGQVLGVFGAIDDAETGFVLTAGEVAGQLAKIGATQPVGTGACVS
>Rv3682 ponA2 TB.seq 4121913:4124342 MW:84637 SEQ ID NO:281
5 MPERLPAAITVLKLAGCCLLASVVATALTFPFAGGLGLMSNRASEVVANGSAQLLEGQVPAVSTMVD
AKGNTIAWLYSQRREFVPSDKIANTMKLAIVSIEDKRFADHSGVDWKGTLTGLAGYASGDLDTRGGS
TLEQQYVKNYQLLVTAQTDAAEKRAAVETTPARKLREIRMALTDKFTFKSEILTRYLNLSFGNNSFG
VQDAAQTYFGINASDLNWQQAALLAGMVQSTSTLNPYTNPDGALARRNVVLDTMENLPGEAEALR
AAKAEPGLVLPQPNELPRGCIAGDRAFFCDYVQEYLSRAGISKEQVATGGYLIRTTLDPEVQAPVKA
AIDKYASPNLAGISSVMSVIKPGKDAHKVLAMASNRKYGLDLEAGETMRPQPFSLVGDGAGSIFKIFT
10 TAAALDMGMGINAQLDVPPRFQAKGLSGGAKGCPKETWCVVNAGNYRGSMNVTDALATSPNTAF
AKLISQVGVGRAVDMAIKLGLRSYANPGTARDYNPDNLSLADFVKRQNLGSFTLGPIELNALELSNV
AATLASGGVWCPPNPIDQLIDRNGNEVAVTTETCDQVVPAGLANTLANAMSKDAVSGSTAAGSAGA
AGWDLPMMSGKTGTTEAHRSAAGFVGFNTNRYAAANYIYDDSSSPTDLCGPLRHCGSGDLYGGNEPS
RTWFAAMKPIANNFGEVQLPPTDPRYVDGAPGSRVPSVAGLDVDAARQLKDAGFQVADQTNQSVN
15 SSAKYGEVVGTSPPSGQTIPGSIVTIQISNGIPPAPPPPLPEDGGPPPPVGSQVVEIPGLPPTIPLAP
PPPPPP
>Rv3721c dnaX DNA polymerase III,[gamma] (dnaZ) and t (dnaX) TB.seq 4164995:4166728
MW:61892 SEQ ID NO:282
20 VALYRKYPASFAEVVGQEHVTAPLSVALDAGRINHAYLFSGPRGCGKTSSARILARSLNCAQGPTA
NPCGVCESCVSLAPNAPGSIDVVELDAASHGGVDDTRELDRAFYAPVQSRVRFVDEAHMVTTA
GFNALLKIVEEPPEHLIFIFATTEPEKVLPTIRSRTTHYPFRLLPPRTMRALLARICEQEGVWDDAVYP
LVIRAGGGSPRDTLSVLDQLLAGAADTHVTYTRALGLLGVTDVALIDDAVDALAACDAAALFGAIESVI
DGGHDPRRFATDLLERFRDLIVLQSVPAASRGVVDAPEDALDRMREQAARIGRATLTRYAEVVQA
GLGEMRGATAPRLLLEVVCARLLLPSASDAESALLQRVERIETRLDMSIPAPQAVPRPSAAAAEPKHQ
25 PAREPRPVLAPTASSEPTVAAVRSMWPTVRDKVRLRSRTTEVMLAGATVRALEDNTLVLTTHESAPL
ARRLSEQRNADVLAEALKDALGVNWRVRCETGEPAASPVGGGANVATAKAVNPAPTANSTQRD
EEEHMLAEAGRGDPSPRRDPEEVALELLQNELGARRIDNA
>Rv3783 - TB.seq 4229255:4230094 MW:32337 SEQ ID NO:283
30 MTFMDAQASFQTSRRTLARVRGDLVDGFRRHFWLHLGWQDIKQRYRRSVLGPFWITATGTTAVA
MGGLYSKLFRLSEHLPPYVTLGLIVWNLINAAILDGAEVFVANGLIKQLPAPLSVHVYRLVWRQMIF
FAHNIVYFVIAIIFPKPWSWADLSFLPALALIFLNCVWVSLCFILATRYRDIGPLFSVVQLLFFMTPII
WNDETLLRQAGRWSSIVELNPLLHYLDIVRAPLLGAHQELRHWLVVLTVVGWMLAAAFAMRQYR
ARVPYVW
>Rv3789 - TB.seq 4235371:4235733 MW:13378 SEQ ID NO:284
35 MRFVVTGGLAGIVDFGLYVVLKYVAGLQVDLSKAISFIVGTITAYLINRRWTFQAEPSTARFVAVMLLY
GITFAVQVGLNHLCLALLHYRAWAIPVAFVIAQGTATVINFIVQRAVIFRIR
>Rv3790 - TB.seq 4235776:4237158 MW:50164 SEQ ID NO:285

MLSVGATTTATRLTGWGRTPASVANL RTPDAEMIVKAVARVAESGGGRGAIARGLGRSYGDNAQN
GGGLVIDMTPLNTIHSIDADTKLVDIDAGVNLDQLMKAALPFGLWVPVLPGTRQVTVGGAIACDIHGK
NHHSAGSFGNHVRSMDLLTADGEIRHLTPTGEDAEFWATVGGNGLTGIIMRATIEMTPTSTAYFIAD
GDVTASLDETIALHSDGSEARYTYSSAWFDAISAPPKLGRAAVSRGRLATVEQLPAKLRSEPLKFDAP
5 QLLTLPDVFPNGLANKYTFGPIGELWYRKSGTYRGKVQNL TQFYHPLDMFGEWNRAYGPAGFLQYQ
FVIPTAEVDEFKKIIGVIAQSGHYSFLNVFKLFGPRNQAPLSFPIPGWNICVDFPIKDGLGKFVSELD RR
VLEFGGRLYTAKDSRTTAETFHAMYPRVDEWISVRRKVDPLRVFASDMARRLELL

>Rv3791 - TB.seq 4237162:4237923 MW:27470 SEQ ID NO:286

MVLDAVGNPQTVLLGGTSEIGLAICERYLHNSAARIVLACL PDDPRREDAAAAMKQAGARSVELIDF
10 DALD TD SHPKMIEAAFSGGDVDVAIVAFGLLGDAEELWQNQRKAVQIAEINYTA AVSVGVLLAEKMR
AQGFQGIIAMSSAAGERVRRANFVYGSTKAGLDGFYLG LSEALREYGVRVLVIRPGQVRTRMSAHLK
EAPLTV DKEYVANLAVTASAKGKELVWAPAAF RYVMMVLRHIPRSIFRKLPI

>Rv3794 embA TB.seq 4243230:4246511 MW:115694 SEQ ID NO:287

VPHDGNERSHRIARLAAV VSGIAGLLCGIVPLLPVNQT TATIFWPQGSTADGNITQITAPLVSGAPRA
15 LDISIPCSAIATLPANGGLVLSTLPAGGVD TGKAGLFVRANQDTVVAFRDSVA AVAARSTIAAGGCS
ALHIWADTGGAGADFMGIPGGAGTLPPEKKPQVGGIFTDLKVG AQPGLSARVDIDTRFITTPGALKKA
VMLLGVLAVLVAMVGLAALDRLSRGRTL RDWLTRYRPRVRVGFASRLADA AVIATLLLWHVIGATSS
DDGYLLTVARVAPKAGYVANYRYFGTTEAPFDWYTSVLAQ LAAVSTAGVWMRLPATLAGIACWLIV
SRFVLRRLGPGPGGLASNRVAVFTAGAVFLSAWL PFNNGLRPEPLIALGVLTWV LVERSIALGR LAP
20 AAVAIIVATLTATLAPQGLIALAPLLTGARAIAQRIRRRRATDGL LAPLAVLAAALSLITVVVFRDQTLATV
AESARIKYKVGPTIAWYQDFLRYF LTVESNVEGSM SRRFAVLVLLFCLFGVLFVLLRRGRVAGLASG
PAWRLIGTTAVGLLLLTFTPTKWAVQFGAFAGLAGVLGAVTAFTFARIGLHSRRNL TLYVTALLFVLA
WATSGINGWFYVGNYPWYDIQPVIA SHPVTSMFLTSLTGLLAAWYHFRMDYAGHTEVKDNRR
NRILASTPLL VAVIMVAGEVGSM AKA AVFRYPLYTTAKANLTALSTGLSSCAMADDVLAEPDPNAGM
25 LQPVPGQA FGPDPGLGGISPVGFKPEGVGEDLKSDPVVSKPGLVNSDASP NKPNAAITDSAGTAGG
KGPVGINGSHAALPFGLDPARTPVMG SYGENNLAATATSAWYQLPPRSPDRPLVVVSAAGAIWSYK
EDGDFIYQSLKLQWGV TGPDGRIQPLGQVFPI DIGPQPAWRNLRFLAWAPPEADVARIVAYDPNL
SPEQWFAFTPPRPVLES LQRLIGSATPVLM DIATAANFPCQRPFSEHLGIAELPQYRILPDHKQTAA
SSNLWQSSSTGGPFLFTQALLRTSTIATYLRGDWYRDWGSVEQYHRLVPADQAPDAVVEEGVITVP
30 GWGRPGPIRALP

>Rv3795 embB TB.seq 4246511:4249804 MW:118023 SEQ ID NO:288

MTQCASRRKSTPNRAILGAFASARGTRWWATIAGLIGFVLSVATPLL PVVQTTAMLDWPQRGQLGSV
TAPLISLT PVDF TATVPCDVVRAMPPAGGVLTAPKQ GKDANLQALFVWVSAQRVDVTD RNVILS
VPREQVTS PQCQRIEVTSTHAGTFANFVGLKDP SGAPLRSGFPDNL RPQIVGVFTDLTG PAPPGLA
35 VSATIDTRFSTRPTTLKLLAIGAIVATVVALIALWRLDQLDGRGSIAQLLLRPFRPASSPGGMRR LIPAS
WRTFTLTDAVWIFGFLWHVIGANSSDDGYILGMARVADHAGYMSNYFRWFGSPEDPFGWYNNLLA
LMTHVSDASLWMRLPDLAAGLVCWLLLSREVL PRLGP AVEASKPAYWAAAMVLLTAWMPFNNGLR

PEGHIALGSLVTYVLIERSMRYSLTPAALAVVTAFTLGVQPTGLIAVAALVAGGRPMLRILVRRHRLV
GTLPLVSPMLAAGTVILT VVFADQTLSTVLEATRVRAKIGPSQAWYTENLRYYYLILPTVDGSLSRFRG
FLITALCLFTAVFIMLRRKRIPSVARGPAWRMLGVIFGTMFFLMFTPTKWVHHFGLFAAVGAAMAALT
TVLVSPSVLRWSRNRMAFLAALFFLLALCWATTNGWWYVSSYGVPFNSAMPKIDGTVSTIFFALFAI
5 AAGYAAWLHFAPRGAGEGRILALTTAPVPIVAGFMAAVFVASMVAGIVRQYPTYSNGWSNVRAFV
GGCGLADDVLVEPDTNAGFMKPLDGDGSGSWGPLGGLGVNPVGFTPNGVPEHTVAEAIMKPNQP
GTDYDWDAPTKLTSPGINGSTVPLPYGLDPAHVPLAGTYTTGAQQQSTLVSAWYLLPKPDDGHPLV
VTAAGKIAGNSVLHGYPGQTVLEYAMPGPAGALVPAGRMVPDDLYGEQPKAWRNLRFAKAMP
ADAVAVRWAEDLSLTPEDWIAVTPRPVPLRSLQEYVGSTQPVLLDWAVGLAFPCQQPMLHANGIA
10 EIPKFRITPDYSAKKLDTDWEDGTNGGLLGITDLLLRAHV MATYLSRDWARDWGSLRKFDTLVDAP
PAQLELGTATRSGLWSPGKIRIGP

>Rv3834c serS seryl-tRNA synthase TB.seq 4307655:4308911 MW:45293 SEQ ID NO:289

VIDLKLLRENPDVRRSQLSRGEDPALVDALLTADAARRAVISTADSLRAEQKAASKSVGGASPEERP
PLLRRAKELAEQVKAEEADEVEAEAAFTAAHLAISNVVDGVPAGGEDDYAVLDVVGEPSPYLENPKD
15 HLELGESLGLIDMQRGAKVSGSRFYFLTGRGALLQLGLLQALKLAVDNGFVPTIPPVLVRPEVMVGT
GFLGAHAEVYRVEGDGLYLVGTSEVPLAGYHSGEILDSRGPLRYAGWSSCFRREAGSHGKDTRG
IIRVHQFDKVEGFVYCTPADAEHEHERLLGWQRQMLARIEVPYRVIDVAAGDLGSSAARKFDCEAWI
PTQGAYRELTSTSNCTTFQARRLATRYRDASGKPKQIAATLNGTLATTRWLVAILENHQRPDGSVRVP
DALVPFVGVEVLEPVA

20 >Rv3907c pcnA polynucleotide polymerase TB.seq 4391631:4393070 MW:53057 SEQ ID NO:290

VPEAVQEADLLTAAVALNRHAALLRELGSVFAAAGHELVLVGGSVRDALLGRLSPDLDFTTDARPE
RVQEIVRPWADAVWDTGIEFGTVGVGKSDHRMEITTFRADSYDRVSRHPEVRFGDCLEGLVRRDF
TTNAMAVRVTATGPGEFLDPLGGLAALRAKVLDTAAPSFSFGDDPLRMLRAARFVSQLGFAVAPR
VRAAIEEMAPQLARISAERVAEELD KLLVGEDPAAGIDLMVQSGMGAVVLPEIGGMRMAIDEHHQHK
25 DVMQHSLSLTLRQAIALEDDGPDVLVLRWAALLHDIGKPTRRHEPDGGVSFHHHEVVGAKMVRKRMR
ALKYSKQMIDDISQLVYLHLRFHGYGDGKWTD SAVRRYVTDAGALLPRLHKLVRADCTTRNKRRAR
LQASYDRLEERIAELAAQEDLDRVRPDLGNQIMAVLDIPAGPQVGEAWRYLKELRLERGPLSTEEA
TTELLSWWKSRGNR

30 A number of embodiments of the invention have been described. Neverthe-
less, it will be understood that various modifications may be made without departing from
the spirit and scope of the invention. Accordingly, other embodiments are within the scope
of the following claims.

WHAT IS CLAIMED IS:

1. A method for identifying a nucleic acid or a polypeptide sequence that may be a target for a drug comprising the following steps:

(a) providing a first nucleic acid or a polypeptide sequence that is known to be a drug target;

(b) providing at least one algorithm selected from the group consisting of a "domain fusion" method, a "phylogenetic profile" method and a "physiologic linkage" method, wherein the algorithm is capable analyzing a functional relationship between nucleic acid or polypeptide sequences; and

(c) comparing the first nucleic acid or the polypeptide drug target sequence to a plurality of sequences using at least one of the algorithms as set forth in step (b) to identify a second sequence that has a functional relationship to the first sequence, thereby identifying a nucleic acid or a polypeptide sequence that may be a target for a drug .

2. A method for identifying a nucleic acid or a polypeptide sequence that may be essential for the growth or viability of an organism comprising the following steps:

(a) providing a first nucleic acid or a polypeptide sequence that is known to be essential for the growth or viability of an organism;

(b) providing at least one algorithm capable analyzing a functional relationship between nucleic acid or polypeptide sequences selected from the group consisting of a "domain fusion" method, a "phylogenetic profile" method and a "physiologic linkage" method; and

(c) comparing the first nucleic acid or the polypeptide sequence to a plurality of sequences using at least one of the algorithms as set forth in step (b) to identify a second sequence that has a functional relationship to the first sequence, thereby identifying a nucleic acid or a polypeptide sequence that may be essential for the growth or viability of an organism.

3. The method of claim 1 or claim 2, wherein the drug is an anti-microbial drug.

4. The method of claim 1 or claim 2, wherein the first nucleic acid or a polypeptide sequence is derived from a pathogen.

5. The method of claim 4, wherein the pathogen is a microorganism.

6. The method of claim 1 or claim 2, wherein the microorganism is *Mycobacterium tuberculosis* (MTB).

7. The method of claim 1 or claim 2, wherein the plurality of sequences used to identify a second sequence comprises a database of the gene sequences of an entire genome of an organism.

8. The method of claim 1 or claim 2, wherein the plurality of sequences used to identify a second sequence comprises a database of the gene sequences derived from a pathogen.

9. The method of claim 1 or claim 2, wherein the "phylogenetic profile" method algorithm comprises

(a) obtaining data, comprising a list of proteins from at least two genomes;

(b) comparing the list of proteins to form a protein phylogenetic profile for each protein, wherein the protein phylogenetic profile indicates the presence or absence of a protein belonging to a particular protein family in each of the at least two genomes based on homology of the proteins; and

(c) grouping the list of proteins based on similar profiles, wherein proteins with similar profiles are indicated to have a functional relationship.

10. The method of claim 9, wherein the phylogenetic profile is in the form of a vector, matrix or phylogenetic tree.

11. The method of claim 9, comprising determining the significance of homology between the proteins by computing a probability (p) value threshold.

12. The method of claim 11, wherein the probability is set with respect to the value $1/NM$, based on the total number of sequence comparisons that are to be performed, wherein N is the number of proteins in the first organism's genome and M in all other genomes.

13. The method of claim 9, wherein the presence or absence is by calculating an evolutionary distance.

14. The method of claim 13, wherein the evolutionary distance is calculated by:

- (a) aligning two sequences from the list of proteins;
- (b) determining an evolution probability process by constructing a conditional probability matrix: $p(aa \rightarrow aa')$, where aa and aa' are any amino acids, said conditional probability matrix being constructed by converting an amino acid substitution matrix from a log odds matrix to said conditional probability matrix;
- (c) accounting for an observed alignment of the constructed conditional probability matrix by taking the product of the conditional probabilities for each aligned pair during the alignment of the two sequences, represented by $P(p) = \prod_n p(aa_n \rightarrow aa'_n)$; and
- (d) determining an evolutionary distance α from powers equation $p' = p^\alpha(aa \rightarrow aa')$, maximizing for P .

15. The method of claim 14, wherein the conditional probability matrix is defined by a Markov process with substitution rates, over a fixed time interval.

16. The method of claim 14, where the conversion from an amino acid substitution matrix to a conditional probability matrix is represented by:

$$P_B(i \rightarrow j) = p(j) 2^{\frac{\text{BLOSUM62}_{ij}}{2}},$$

where BLOSUM62 is an amino acid substitution matrix, and $P(i \rightarrow j)$ is the probability that amino acid i is replaced by amino acid j through point mutations according to BLOSUM62 scores.

17. The method of claim 16, where P_j 's are the abundances of amino acid j and are computed by solving a plurality of linear equations given by the normalization condition that:

$$\sum_i P_B(i \rightarrow j) = 1.$$

18. The method of claim 1 or claim 2, wherein the "physiologic linkage" method algorithm identifies proteins and nucleic acids that participate in a common functional pathway.

19. The method of claim 1 or claim 2, wherein the "physiologic linkage" method algorithm comprises identifies proteins and nucleic acids that participate in the synthesis of a common structural complex.

20. The method of claim 1 or claim 2, wherein the "physiologic linkage" method algorithm comprises identifies proteins and nucleic acids that participate in a common metabolic pathway.

21. The method of claim 1 or claim 2, wherein the "domain fusion" method algorithm comprises

- (a) aligning a first primary amino acid sequence of multiple distinct non-homologous polypeptides to second primary amino acid sequence of a plurality of proteins; and
- (b) for any alignment found between the first primary amino acid sequences of all of such multiple distinct non-homologous polypeptides and at least one protein of the second primary amino acid sequences, outputting an indication identifying the aligned second primary amino acid sequence as an indication of a functional link between the aligned first and second polypeptide sequences.

22. The method of claim 21, wherein the aligning is performed by an algorithm selected from the group consisting of a Smith-Waterman algorithm, Needleman-Wunsch algorithm, a BLAST algorithm, a FASTA algorithm, and a PSI-BLAST algorithm.

5 23. The method of claim 21, wherein the multiple distinct non-homologous polypeptides are obtained by translating a nucleic acid sequence from a genome database.

10 24. The method of claim 21, wherein the plurality of proteins have a known function.

25. The method of claim 21, wherein at least one of the multiple distinct non-homologous polypeptides has a known function.

15 26. The method of claim 21, wherein at least one of the multiple distinct non-homologous polypeptides has an unknown function.

20 27. The method of claim 21, wherein the alignment is based on the degree of homology of the multiple distinct non-homologous polypeptides to the plurality of proteins.

28. The method of claim 21, further comprising determining the significance of the aligned and identified second primary amino acid sequence by computing a probability (p) value threshold.

25 29. The method of claim 28, wherein the probability threshold is set with respect to the value $1/NM$, based on the total number of sequence comparisons that are to be performed, wherein N is the number of proteins in a first organism's genome and M in all other genomes.

30 30. The method of claim 21, further comprising filtering excessive functional links between one first primary amino acid sequence of multiple distinct non-

homologous polypeptides and an excessive number of other distinct non-homologous polypeptides for any alignment found between the first primary amino acid sequences of the distinct non-homologous polypeptides and at least one of the second primary amino acid sequences of the plurality of proteins.

5

31. A computer program product, stored on a computer-readable medium, for identifying a nucleic acid or a polypeptide sequence that may be a target for a drug, the computer program product comprising instructions for causing a computer system to be capable of:

10

(a) inputting a first nucleic acid or a polypeptide sequence that is known to be a drug target;

15

(b) accessing at least one algorithm capable analyzing a functional relationship between nucleic acid or polypeptide sequences selected from the group consisting of a "domain fusion" method, a "phylogenetic profile" method and a "physiologic linkage" method; and

20

(c) comparing the first nucleic acid or the polypeptide drug target sequence to a plurality of sequences using at least one of the algorithms set forth in step (b) to identify a second sequence that has a functional relationship to the first sequence and generating an output identifying a nucleic acid or a polypeptide sequence that may be a target for a drug .

25

32. A computer program product, stored on a computer-readable medium, for identifying a nucleic acid or a polypeptide sequence that may be essential for the growth or viability of an organism, the computer program product comprising instructions for causing a computer system to be capable of:

(a) providing a first nucleic acid or a polypeptide sequence that is known to be essential for the growth or viability of an organism;

30

(b) accessing at least one algorithm capable analyzing a functional relationship between nucleic acid or polypeptide sequences selected from the group consisting of a "domain fusion" method, a "phylogenetic profile" method and a "physiologic linkage" method; and

(c) comparing the first nucleic acid or the polypeptide sequence to a plurality of sequences using at least one of the algorithms set forth in step (b) to identify a second sequence that has a functional relationship to the first sequence and generating an output identifying a nucleic acid or a polypeptide sequence that may be essential for the growth or viability of an organism.

5

33. A computer system, comprising:

(a) a processor; and

(b) a computer program product as set forth in claim 31 or claim 32.

10

Figure 1

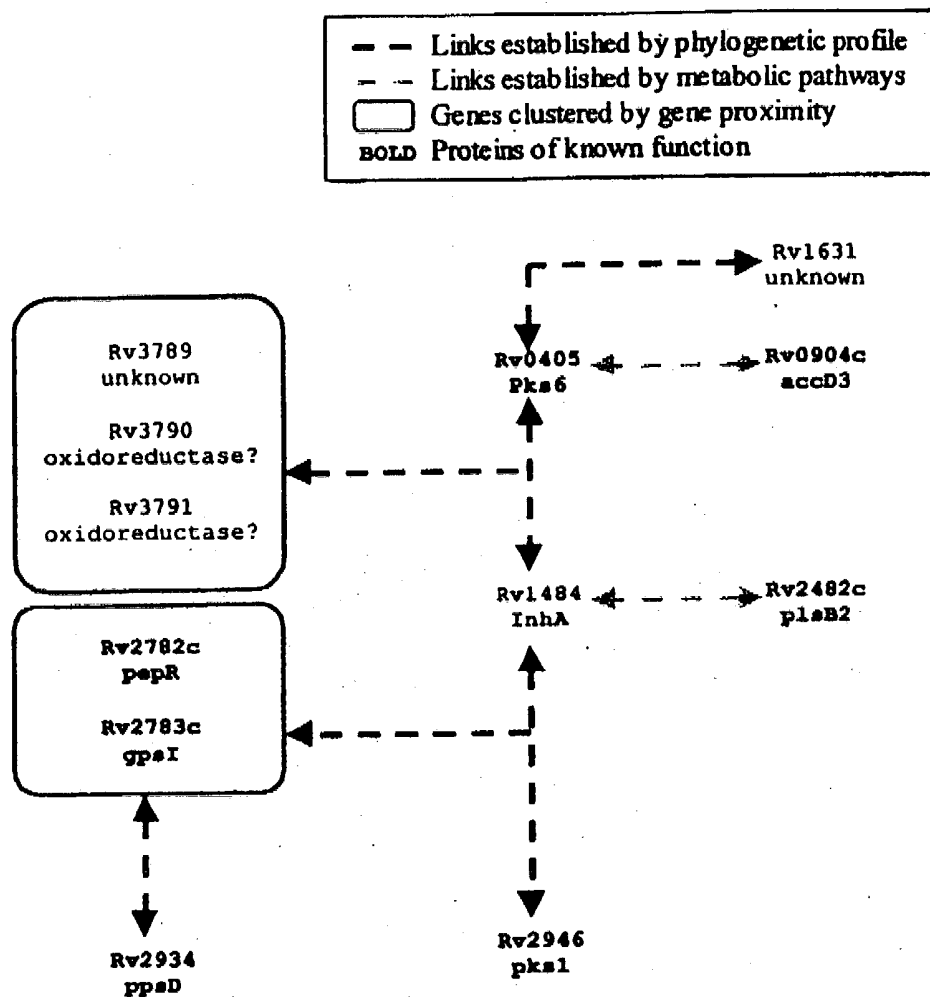


Figure 2

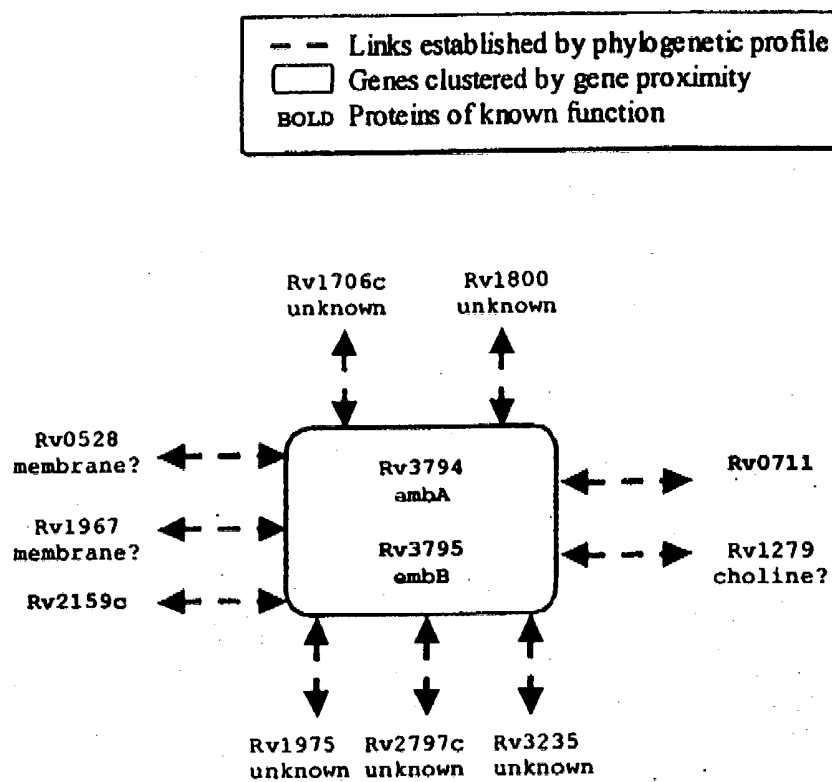


Figure 3

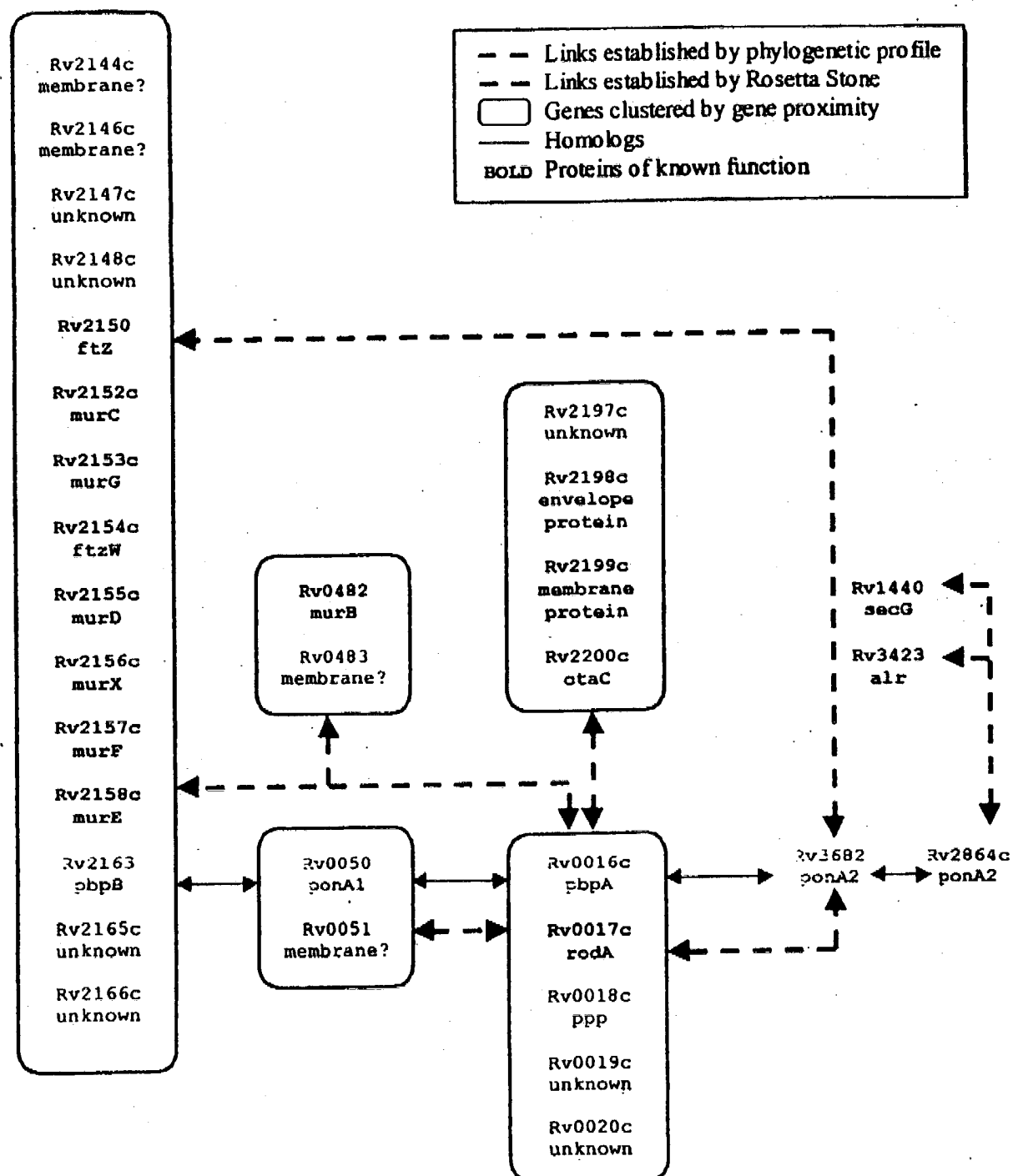


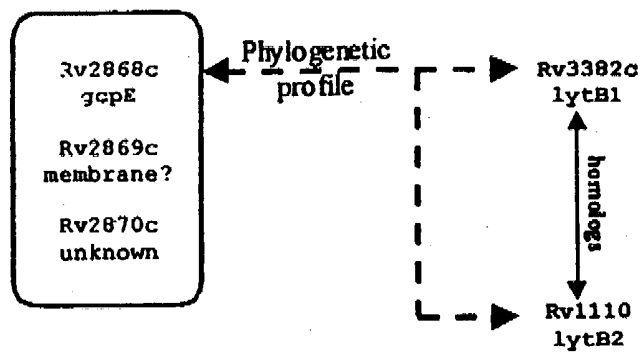
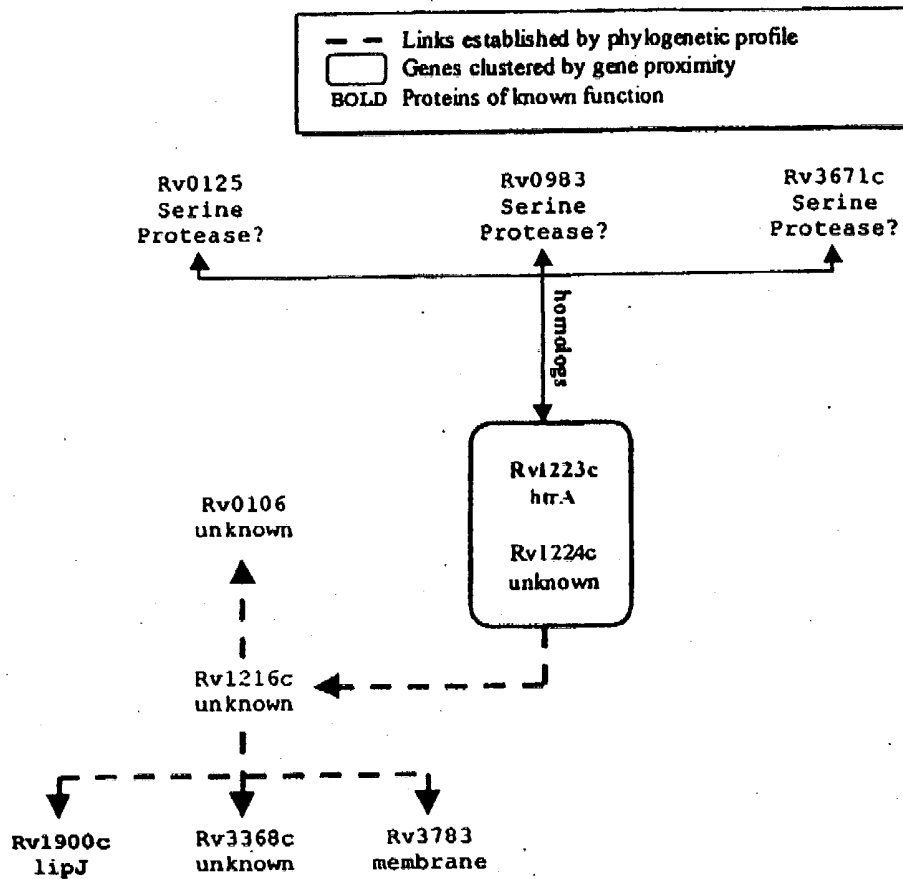
Figure 4

Figure 5



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/31152

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 19/00

US CL : 702/19

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 702/19

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

STN ON LINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MARCOTTE et al " A COMBINED ALGORITHM FOR GENOMEWIDE PREDICTION OF PROTEIN FUNCTION" NATURE GB, 04 November 1999, vol. 402, no. 6757, pages 83-86	1-33
X	PELLEGRINI M et al. "Assigning protein functions by comparative genome analysis: protein phylogenetic profiles". April 1999, vol. 96, pages 4285-4288; the whole document	1-33
X	ENRIGHT A J et al. Protein interaction maps for complete genomes based on gene fusion elements. Nature, 04 November 1999, Vol. 402, pages 86-90, the whole document	1-33
X	TATUSOV R L et al. A Genomic perspective on protein families. Science, 24 October 1997, vol. 278, pages 631-637, the whole document.	1-33

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
B earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

28 FEBRUARY 2001

Date of mailing of the international search report

11 APR 2001

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

MICHAEL BORIN

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/31152

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MARÇOTTE E et al. Detecting protein function and protein-protein interactions from genome sequences. Science, 30 July 1999, vol. 285, pages 751-753, the whole document	1-33
A	EISEN M et al. Cluster analysis and display of genome-wide expression patterns. Proceedings of Natl. Acad. Sci., USA, December 1998, vol. 95, pages 14683-14868.	1-33
X,P	WO 00/45322 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA), 03 August 2000, claims 1-77.	1-33